

**UNIVERSIDAD CATÓLICA
DE SANTIAGO DE GUAYAQUIL**

**FACULTAD DE ECONOMÍA Y EMPRESA
CARRERA DE NEGOCIOS INTERNACIONALES**

TÍTULO:

Creación de un modelo clasificador para oportunidades de captación en el exterior por Machine Learning de la empresa EDEX Education Group.

AUTORES:

Cuesta Asanza, Ximena Margarita

Duran Encalada, Eduardo Javier

**Trabajo de integración curricular previo a la obtención de título de
LICENCIADO EN NEGOCIOS INTERNACIONALES**

TUTOR:

Ing. Carrera Buri, Félix Miguel, Mgs.

Guayaquil, Ecuador

16 de febrero del 2024



UNIVERSIDAD CATÓLICA
DE SANTIAGO DE GUAYAQUIL

FACULTAD DE ECONOMÍA Y EMPRESA
CARRERA NEGOCIOS INTERNACIONALES

CERTIFICACIÓN

Certificamos que el presente trabajo de titulación fue realizado en su totalidad por **Cuesta Asanza, Ximena Margarita y Duran Encalada, Eduardo Javier** como requerimiento para la obtención del título de **Licenciado en Negocios Internacionales**.

TUTOR

f. _____

Ing. Carrera Buri, Félix Miguel, Mgs.

DIRECTORA DE LA CARRERA

f. _____

Ing. Hurtado Cevallos, Gabriela Elizabeth, Mgs

Guayaquil, a los 16 días del mes de febrero del año 2024



UNIVERSIDAD CATÓLICA
DE SANTIAGO DE GUAYAQUIL

FACULTAD DE ECONOMÍA Y EMPRESA
CARRERA NEGOCIOS INTERNACIONALES
DECLARACIÓN DE RESPONSABILIDAD

Nosotros, **Cuesta Asanza, Ximena Margarita**
Duran Encalada, Eduardo Javier.

DECLARAMOS QUE:

El Trabajo de Titulación, **Creación de un modelo clasificador para oportunidades de captación en el exterior por Machine Learning de la empresa EDEX Education Group**, previo a la obtención del título de **Licenciados en Negocios Internacionales**, ha sido desarrollado respetando derechos intelectuales de terceros conforme las citas que constan en el documento, cuyas fuentes se incorporan en las referencias o bibliografías. Consecuentemente este trabajo es de mi total autoría.

En virtud de esta declaración, me responsabilizo del contenido, veracidad y alcance del Trabajo de Titulación referido.

Guayaquil, a los 16 días del mes de febrero del año 2024

LOS AUTORES:

f.

Cuesta Asanza, Ximena Margarita

f.

Duran Encalada, Eduardo Javier



UNIVERSIDAD CATÓLICA
DE SANTIAGO DE GUAYAQUIL

FACULTAD DE ECONOMÍA Y EMPRESA
CARRERA NEGOCIOS INTERNACIONALES

AUTORIZACIÓN

Nosotros, Cuesta Asanza, Ximena Margarita
Duran Encalada, Eduardo Javier

Autorizamos a la Universidad Católica de Santiago de Guayaquil a la **publicación** en la biblioteca de la institución del Trabajo de Titulación, **Creación de un modelo clasificador para oportunidades de captación en el exterior por Machine Learning de la empresa EDEX Education Group**, cuyo contenido, ideas y criterios son de nuestra exclusiva responsabilidad y total autoría.

Guayaquil, a los 16 días del mes de febrero del año 2024

LOS AUTORES:

f. _____

Cuesta Asanza, Ximena Margarita

f. _____

Duran Encalada, Eduardo Javier



**UNIVERSIDAD CATÓLICA
DE SANTIAGO DE GUAYAQUIL
FACULTAD DE ECONOMÍA Y EMPRESA
CARRERA DE NEGOCIOS INTERNACIONALES**

CERTIFICADO DE ANÁLISIS
magister

Tesis Final XIMENA CUESTA ASANZA - EDUARDO DURAN ENCALADA

3% Textos sospechosos

2% Similitudes
0% similitudes entre comillas
< 1% entre las fuentes mencionadas
< 1% idioma no reconocido

Nombre del documento: Tesis Final XIMENA CUESTA ASANZA - EDUARDO DURAN ENCALADA.docx
ID del documento: d5d5166124caab64324b7c6a845ccf98b18aa31
Tamaño del documento original: 3,36 MB

Depositante: Felix Miguel Carrera Buri
Fecha de depósito: 26/1/2024
Tipo de carga: interface
fecha de fin de análisis: 26/1/2024

Número de palabras: 21.960
Número de caracteres: 144.850

Ubicación de las similitudes en el documento:

Fuentes principales detectadas

Nº	Descripciones	Similitudes	Ubicaciones	Datos adicionales
1	repositorio.ucsg.edu.ec 15 fuentes similares	1%		Palabras idénticas: 1% (218 palabras)
2	repositorio.ucsg.edu.ec 13 fuentes similares	1%		Palabras idénticas: 1% (235 palabras)
3	repositorio.ucsg.edu.ec 16 fuentes similares	1%		Palabras idénticas: 1% (222 palabras)
4	TESIS.docx text: #a7c0b El documento proviene de mi grupo 9 fuentes similares	< 1%		Palabras idénticas: < 1% (193 palabras)
5	repositorio.ucsg.edu.ec 8 fuentes similares	< 1%		Palabras idénticas: < 1% (120 palabras)

f. _____

Ing. Carrera Buri, Félix Miguel, Mgs

TUTOR

AGRADECIMIENTO

Quisiera expresar mi sincero agradecimiento a mis padres que han estado conmigo en cada logro que he conseguido y me han apoyado en todos los proyectos que he hecho, a mis abuelos Paco y Luisita por ser mi fuente de sabiduría desde que soy muy pequeña y en especial a mi tío Santiago y mi tía Gloria, que han sido un apoyo fundamental en mi vida personal y académica.

Todos son ejemplo de dedicación y sacrificio y estoy profundamente agradecida por su guía en cada paso que doy.

-Ximena Margarita Cuesta Asanza

AGRADECIMIENTO

A través de estas líneas, quiero expresar mi profundo agradecimiento hacia mi mamá, por su apoyo, amor y paciencia que han sido mi mayor fortaleza. Gracias por ser mi inspiración en los momentos difíciles, ya que sin ti no estuviera aquí.

Este logro no solo es mío, sino también tuyo, gracias por ser mi mayor admiradora, te amo.

Tampoco quería dejar pasar por alto a usted Cesar, gracias por ser ese amigo y apoyo en este camino y una de las mejores cosas que pudo pasarme en el transcurso de la universidad es conocer a mi compañera Ximena que sin el apoyo y amistad de ella no hubiera llegado a la mitad de las clases.

-Eduardo Javier Duran Encalada

DEDICATORIA

A mis padres, por su apoyo incondicional.

-Ximena Margarita Cuesta Asanza

DEDICATORIA

A mi familia que me apoyo en todo este camino.

-Eduardo Javier Duran Encalad



**UNIVERSIDAD CATÓLICA
DE SANTIAGO DE GUAYAQUIL
FACULTAD DE ECONOMÍA Y EMPRESA
CARRERA DE NEGOCIOS INTERNACIONALES**

TRIBUNAL DE SUSTENTACIÓN

f. _____

Ing. Hurtado Cevallos, Gabriela Elizabeth, Mgs
DIRECTORA DE CARRERA

f. _____

Ing.
COORDINADOR DEL ÁREA

f. _____

(NOMBRES Y APELLIDOS)
OPONENTE



**UNIVERSIDAD CATÓLICA
DE SANTIAGO DE GUAYAQUIL
FACULTAD DE ECONOMÍA Y EMPRESA
CARRERA DE NEGOCIOS INTERNACIONALES**

CALIFICACIÓN

f. _____

Cuesta Asanza, Ximena Margarita

f. _____

Duran Encalada, Eduardo Javier

Guayaquil, a los 16 días del mes de febrero del año 2024

ÍNDICE DE CONTENIDO

I.	INTRODUCCIÓN.....	2
1.	Antecedentes	2
2.	Problemática.....	6
3.	Justificación	8
4.	Alcance	13
5.	Objetivos	15
5.1.	Objetivo General	15
5.2.	Objetivos Específicos	15
6.	Preguntas de investigación planteadas.....	15
II.	CAPÍTULO I: MARCO TEÓRICO	16
1.	Globalización	16
2.	Mercado Laboral.....	17
3.	La importancia de los idiomas	18
4.	Importancia de la formación académica de alta calidad.....	19
5.	Importancia de la experiencia laboral.....	20
6.	Empresa EDEX EDUCATION GROUP	21
7.	Inteligencia Artificial	23
8.	Inteligencia artificial en el entorno económico y mercado	25
9.	Machine Learning	27
10.	Tipos de Machine Learning.....	29
11.	Regresión Lineal.....	33
12.	KMedias para la separación de grupos de importancia	34
14.	Random Forest	37
15.	Clustering.....	38
III.	CAPITULO III: MARCO LEGAL.....	40
IV.	CAPÍTULO IV: METODOLOGIA.....	41
1.	Etapa 1: Implementación de los algoritmos K Medias para la toma de decisiones 41	
2.	Etapa 2: Pasos a desarrollar.....	41
V.	CAPÍTULO V: RESULTADOS.....	47
VI.	CAPÍTULO VI: CONCLUSIÓN	55
1.	Conclusión.....	55
VII.	CAPÍTULO VII: BIBLIOGRAFÍA.....	57
	ANEXOS	64

ÍNDICE DE TABLAS

Tabla 1 Nivel de importancia.....	51
Tabla 2 Matriz de Confusión.....	52
Tabla 3 Tabla de frecuencia por nivel educativo	52

ÍNDICE DE FIGURAS

Figura 1 Funcionamiento de Machine Learning	28
Figura 2 Modelo de aprendizaje supervisado	29
Figura 3 Algoritmo Machine Learning no supervisado	31
Figura 4 Algoritmo Machine Learning por refuerzo	32
Figura 5 Esquema del árbol de decisión	36
Figura 6 Gráfico de clústeres	47
Figura 7 Gráfico de clústeres	48
Figura 8 Gráfico de clústeres conglomerados	49
Figura 9 Gráfico elíptico tipo euclidiana	50
Figura 10 Gráfico de aplicación de k-means en los clústeres	51
Figura 11 Predicciones a través de K-means y RandomForest	53
Figura 12 Frecuencia tipo de cliente	53
Figura 13 Tipo de cliente por nivel educativo	54

RESUMEN

Este trabajo tiene como objetivo crear un modelo clasificador para la captación de clientes en el exterior de la empresa EDEX Education Group basado en Machine Learning para identificar y evaluar las oportunidades de captación en el ámbito internacional, EDEX Education Group es una empresa que se dedica a la consultoría de estudios en el extranjero, cuenta con oficinas en distintos países pero debido a la pandemia laboran en algunos países a través de video llamadas, en Ecuador cuenta con oficinas las ciudades de Quito, Machala y Guayaquil. La globalización nos ha brindado la oportunidad de poder tomar cursos de idiomas, certificaciones y carreras universitarias en cualquier país del mundo y este proyecto es un enfoque que puede ayudar a mejorar significativamente la eficiencia del proceso de captación y así tomar decisiones más exactas a la hora de trabajar con cada cliente y sus características a través de las consultorías personalizadas, empleando el algoritmo de Random Forest en RStudio.

Palabras claves: Machine Learning, estudios en el extranjero, globalización, RandomForest.

ABSTRACT

This work aims to create a classifier model for the recruitment of customers abroad of the company EDEX Education Group based on Machine Learning to identify and evaluate recruitment opportunities in the international arena, EDEX Education Group is a company dedicated to consulting studies abroad, has offices in different countries but due to the pandemic work in some countries through video calls, in Ecuador they have offices in the cities of Quito, Machala and Guayaquil. Globalization has given us the opportunity to take language courses, certifications and university careers in any country in the world and this project is an approach that can help to significantly improve the efficiency of the recruitment process and thus make more accurate decisions when working with each client and their characteristics through personalized consulting, using the RandomForest algorithm in RStudio.

Keywords: Machine Learning, study abroad, globalization, RandomForest.

RÉSUMÉ

Ce travail vise à créer un modèle de classificateur pour le recrutement de clients à l'étranger de la société EDEX Education Group basé sur l'apprentissage automatique pour identifier et évaluer les opportunités de recrutement dans l'arène internationale, EDEX Education Group est une société dédiée à la consultation des études à l'étranger, a des bureaux dans différents pays, mais en raison de la pandémie de travail dans certains pays par le biais d'appels vidéo, en Équateur a des bureaux dans les villes de Quito, Machala et Guayaquil. La mondialisation nous a donné la possibilité de suivre des cours de langue, d'obtenir des certifications et des diplômes universitaires dans n'importe quel pays du monde et ce projet est une approche qui peut aider à améliorer de manière significative l'efficacité du processus de recrutement et ainsi prendre des décisions plus précises lorsqu'il s'agit de travailler avec chaque client et ses caractéristiques à travers des conseils personnalisés, en utilisant l'algorithme RandomForest dans RStudio.

Mots-clés: Apprentissage automatique, études à l'étranger, globalisation, RandomForest.

I. INTRODUCCIÓN

EDEX Education Group es una empresa dedicada a la experimentación de un entorno laboral y educativo que es necesario en un entorno globalizado. Se ofrece clases en línea y fomenta los viajes al exterior para estudiar y experimentar la educación en el extranjero. La empresa formada en 2014 y que arribó al Ecuador en 2021, representa una oportunidad para que los clientes puedan cumplir sus sueños de estudiar en el extranjero por medio del total asesoramiento en el proceso de inicio de la aventura. Los principios de la empresa son la libertad de elección, la libre adscripción y la participación para construir un espacio educativo que sea innovador y que las personas puedan estar integradas a las necesidades actuales. A continuación, se explicará una forma eficiente de fortalecer la experiencia de EDEX Education Group mediante la intervención de los sistemas estadísticos mediante la herramienta de Machine Learning, por medio de la técnica de Random Forest.

1. Antecedentes

Los cambios en el entorno laboral, la globalización y el desarrollo de nuevos tipos de trabajo, han convertido el aprendizaje de los idiomas en un aspecto fundamental en el desarrollo y crecimiento profesional de las personas. Es por ello que, marcas como Open English, English Live, Babbel, Mondly, Duolingo, Master English, Preply, entre otras, han desarrollado un modelo de negocios enfocado al desarrollo profesional de las personas en el aprendizaje de nuevos lenguajes, siendo el inglés el más importante de estos. El conocimiento de los nuevos idiomas permite que las personas se puedan relacionar entre sí, sin importar que provengan de otros países, aumentando la posibilidad de desarrollar un modelo de comunicación donde se mejoran las habilidades de aprendizaje y desarrollo profesional (Moreno, 2015).

Es por ello que ante el crecimiento de la necesidad de la población en aprender idiomas ha surgido la empresa en el Ecuador EDEX Education Group Guayaquil, formada el 29 de enero de 2021 en el Ecuador, cuya sede principal se encuentra en Quito y posee Escuelas de Lenguaje en el país. La empresa posee ingresos netos de 4037.04% y un crecimiento de 909.59% (EMIS, 2023). La empresa se ha consolidado en todo el mundo

como una Empresa de Consultoría de estudios en el extranjero con los cuales se consolida y fundamenta la necesidad de estudiar idiomas para las personas en general, mediante cursos o por medio de un programa universitario donde el asesoramiento personalizado es fundamental para el crecimiento profesional de la persona.

Para ello, la empresa ofrece asesorías personalizadas que fomentan el estudio del inglés y la educación para el nivel superior en el extranjero y esta se forma por medio de consultores académicos, profesionales con gran capacidad internacional y reconocimiento en el mercado. La empresa ofrece facilidades de acceso para todos los estudiantes, a fin de que estos puedan aumentar su capacidad de estudio y desarrollo académico en el extranjero con los mejores profesionales y expandir sus oportunidades en el ámbito profesional a nivel mundial.

EDEX Education Group está conformado por un equipo de trabajo donde la juventud y la creatividad son más importantes que la experiencia, donde los jóvenes que han experimentado la vida en el extranjero son los responsables de desarrollar un enfoque de trabajo efectivo, combinando aspectos como la vida, el estudio y el trabajo en otros países, durante un periodo que va desde los seis meses hasta los dos años. Para ello, todos los profesionales de la empresa otorgan un reconocimiento respectivo para fomentar el asesoramiento a todos los estudiantes que elijan la empresa para aprender idiomas. Todo el equipo de trabajo que ha formado la empresa posee cualidades como el compromiso, la responsabilidad, el acompañamiento y la calidad. Como compromiso, todos los colaboradores y miembros del equipo de la empresa se comprometen a dar el 100% de todos los beneficios para satisfacción de las metas de los estudiantes y alcanzar sus sueños, gracias a las instituciones de alta calidad.

La marca está comprometida con la responsabilidad para acatar las normativas de las distintas embajadas e instituciones con las cuales se espera que cubran con las necesidades y expectativas de todos los clientes, permitiéndole a la empresa destacar por su calidad educativa. La empresa también se compromete con el acompañamiento y total seguimiento de todos sus clientes, permitiéndoles ir en todo momento de su proceso académico y que puedan lograr la respectiva seguridad del crecimiento educativo del estudiante. La calidad es su sello de garantía institucional, por medio de la cual se ha logrado un importante reconocimiento académico internacional, garantizando opciones

claras de estudio para todas las personas interesadas en el proceso educativo y que este mantenga un alto desarrollo profesional en el extranjero.

Los beneficios que EDEX otorga a todos sus estudiantes es la consultoría personalizada, la elección del destino que el estudiante prefiera, conociendo aspectos como su programa educativo según las necesidades educativas que esta posea. Además, EDEX ofrece resolver problemas como la asesoría y desarrollo en el trámite de las visas estudiantiles y los tiques para los viajes. EDEX se encarga del alojamiento en el extranjero y la recepción incluso desde el aeropuerto, para evitar que los estudiantes puedan extraviarse o sentirse inseguros en un entorno distinto. EDEX ofrece seguros médicos en el viaje internacional, así como becas de estudio superior y de idiomas para los usuarios y siempre otorgarán charlas de recomendación antes del viaje (EDEX Education Group, 2023).

EDEX Education Group ofrece una estrategia de desarrollo comercial que compite con empresas que ofrecen no solo el servicio académico de viajes para estudiantes. Puesto que, por supuesto, los mayores competidores en el mercado son las plataformas digitales y de web que sean pagadas o gratuitas, las cuales ofrezcan la enseñanza académica de inglés para las personas. Entre las más reconocidas se encuentran Open English, Babbel, Platzi English, Duolingo y Mondly. Open English es la marca más popular con el 58% de todos los clientes gracias a su publicidad emitida en redes sociales y en televisión de paga. La empresa que se desarrolló desde Estados Unidos en 2008 para beneficio de la comunidad latina ha sido desarrollada para ofrecer clases personalizadas a sus usuarios, con la propuesta de ofrecer profesores norteamericanos en vivo, la convierten en el sitio más rentable y con mejores beneficios para los usuarios. Los precios son relativamente bajos, a un valor desde \$75.00 al mes, garantizan que los usuarios aprenderían el idioma inglés con ayuda de profesores nativos.

Como competencia directa se encuentran otras marcas como EF Education First, siendo una marca en el Ecuador que ofrece el servicio de viajes para sus clientes. Se enfoca en la diversión y el estudio, y con más de 40 años de experiencia, ofrecen que el aprendizaje del idioma inglés sea una experiencia de alto impacto y crecimiento profesional para todos sus clientes. Además, expone de forma muy clara que cuenta con más de 2 millones de estudiantes que se han beneficiado del servicio. En su propio sitio

web, ofrece la posibilidad a los propios clientes de elegir al grupo según la edad que se tenga, dando como última pauta, la opción de adultos profesionales (EF Ecuador, 2023).

En la actualidad, EDEX posee registrado más de 155 mil estudiantes en todos los 12 países y 65 ciudades que funciona y mantiene interacción con otros institutos educativos del extranjero. Los precios varían según el país, pero en promedio los costos van desde \$4680 dólares por el servicio de viaje de un año donde se trabajará y se estudiará el idioma inglés. Además, se podrá recibir cursos online hasta seis meses antes de viajar a la ciudad de Sídney, Australia (EDEX Education Group, 2023). Los clientes son los que pueden recibir beneficios como los cursos por un lapso mínimo de nueve meses, asesoría, acompañamiento y tips para viajar y conseguir el trabajo. Sin embargo, sus puntos negativos son el no ofrecer beneficios como pasajes de avión, alojamiento, alimentación, seguro médico, materiales, registro de matrícula, costos de visado y la búsqueda del trabajo. Lo cual puede considerarse para el usuario como que la empresa únicamente gestiona el un nombre de marca y las clases de inglés, pero no otorga beneficios que puedan reducir el costo final del servicio. EDEX Education Group además otorga otros beneficios como la enseñanza no solo del idioma inglés, puesto que también se enfoca en ofrecer idiomas como el francés, el alemán y el italiano, gracias a su enfoque de alianzas con centros académicos similares para fomentar el proceso educativo.

El entorno actual invita a la sociedad a desarrollar lo más posible sus cualidades cognitivas y desarrollo intelectual, siendo el aprendizaje de nuevos idiomas necesario para mejorar su calidad de vida, puesto que la globalización así lo determina. Como política educacional en el Ecuador se manifiesta la necesidad de modificar y transformar el proceso educativo de las personas mediante los entornos institucionales, actores y factores sociales que buscan adaptarse y transformarse de forma periódica en base a las nuevas necesidades del entorno educativo. Su objetivo es ofrecer el máximo potencial educativo por medio de las instituciones privadas y públicas para los niños, niñas, adolescentes y jóvenes en el Ecuador (Segovia, 2022).

El concepto de inteligencia artificial consiste en el desarrollo y utilización de algoritmos con la finalidad de desarrollar una simulación del pensamiento humano, imitando así los procesos de desarrollo cognitivo. El desarrollo de la inteligencia necesita de la implementación de los sistemas computacionales, datos y algoritmos avanzados de Inteligencia Artificial para simular el comportamiento del ser humano. El desarrollo de

la inteligencia artificial se generó desde el año 1955 con la idea de John McCarthy, para desarrollar una simulación del pensamiento del ser humano en los sistemas informáticos de forma autónoma mediante el uso del aprendizaje profundo, el análisis predictivo y el estudio prescriptivo (Ponce, Torres, Quezada, & Silva, 2014). La importancia actual de la inteligencia artificial es que permite mejorar la toma de decisiones que genera el ser humano gracias al proceso de aprendizaje automático, mejorando los resultados y los tiempos de respuesta. Actualmente, la inteligencia artificial es utilizada en algunos aspectos de la vida cotidiana como el uso de las herramientas en la medicina para determinar los posibles problemas médicos de un paciente, la detección del fraude financiero, determinación el desarrollo y cambios en el comportamiento del mercado. Así como ser de ayuda para determinar el perfil del consumidor (NETAPP, 2023).

El Machine Learning por otro lado consiste en una rama de la inteligencia artificial que ha sido desarrollada con la finalidad de fomentar el desarrollo y proceso de aprendizaje del sistema computacional, de la ciencia informática y de la inteligencia artificial. Para el funcionamiento del Machine Learning se ha generado el uso de algoritmos y modelos estadísticos los cuales son utilizados para el desarrollo y crecimiento de los modelos estadísticos con la finalidad de generar tareas sin instrucciones explícitas, evitando que estas se basen en patrones e inferencias. Su importancia radica en que son utilizados principalmente para determinar resultados en base a los ingresos de datos que sean simples o complejos, resolviéndolos y obteniendo un resultado. El concepto de Machine Learning se enfoca principalmente en establecer el comportamiento de los clientes, trabajadores y modelos financieros para beneficio de las empresas (Amazon, 2023).

2. Problemática

La deserción de los clientes es un factor que siempre ha afectado a las empresas, la cual es causada cuando el consumidor toma la decisión de dejar de utilizar un determinado producto o servicio, sea por las razones que tenga personales o por el descontento con el producto o servicio. En algunos casos, la deserción de los clientes suele darse simplemente porque no desea más recibir un servicio o consumir un determinado producto. Esta deserción puede generarse de forma voluntaria o involuntaria, dependiendo de los cambios que realice la empresa y afecten al cliente o los cambios generados por el mercado. Incluso, el personal, el tipo de servicio y los productos que

ofrezca la empresa podrían generar cambios en el comportamiento del consumidor (López, 2023).

La pérdida de los clientes trae como consecuencia el encarecer y generar que el servicio se desarrolle de forma más lenta, siendo los casos más tradicionales el descontento o la insatisfacción de los clientes. Lo que puede causar que la mala experiencia o su descontento genere la deserción del resto de consumidores a causa de la mala publicidad de boca en boca, siendo esta la publicidad más efectiva para una empresa. Por lo cual, evitar la pérdida de los clientes es una labor necesaria que la empresa deberá de desarrollar para mantener una relación duradera con el mismo, considerando la premisa que obtener un nuevo cliente es más costoso y complejo que mantener al cliente ya cautivo (Valera, Vásquez, & Iglesias, 2009).

El problema de investigación sobre la empresa EDEX Education Group está enfocado en la existencia de gran parte de los estudiantes de la empresa, los cuales realizan el pago del 30% del costo total del servicio, evitando viajar y abandonando el curso, generando pérdidas para la empresa tanto por la pérdida del cliente, por los costos que generó durante el tiempo que estuvo vigente su participación y por los errores en las proyecciones financieras de la empresa para generar estrategias comerciales y planes de proyecto (EDEX Education Group, 2023). Por ende, se ha planteado la necesidad de utilizar el concepto de Machine Learning para analizar a todos los clientes registrados por la empresa EDEX Education Group, y mediante el análisis de su perfil de consumo, establecer el número de clientes que posiblemente puedan desertar.

Con esta información, la empresa desarrollará nuevas estrategias de mercado como promociones, descuentos y estrategias que eviten reducir hasta en un 40% la deserción de los clientes, a fin de aplicar los modelos predictivos para que la deserción de clientes se reduzca significativamente en la empresa para mantener eficiente su rentabilidad (Actuarial Asesoramiento Estratégico, 2020). Es decir que, por medio de la Machine Learning, se buscará mejorar la experiencia de consumo de los consumidores que forman parte de EDEX Education Group, al mismo tiempo que se logrará desarrollar estrategias de mercado que capten la atención de nuevos consumidores, manteniendo cautivo y fidelizando a los clientes de la empresa, para lograr una mayor rentabilidad.

3. Justificación

El uso de las nuevas tecnologías se ha vuelto una constante en un entorno globalizado, puesto que es gracias a ellas que se han generado los mayores cambios en el desarrollo comercial por medio de la integración de los avances y de las herramientas técnicas en todo el mundo, causando aspectos como el aprovechamiento de las materias primas para lograr un mejor resultado. Todos los cambios en el mercado son los responsables de generar la necesidad que la población se deba preparar y especializar de mejor forma para mantenerse competitivo en el mundo laboral, considerando dos aspectos fundamentales como el uso de las herramientas tecnológicas y el conocimiento de los idiomas, principalmente el inglés (Pino, 2016).

Los conceptos de Machine Learning fueron desarrolladas por primera vez como una idea de análisis de la información desde 1959, siendo este un concepto de emulación del proceso de aprendizaje de los seres humanos generado mediante el uso de computadores, permitiendo procesar un gran volumen de datos de manera rápida. Las Machines Learning representan la transformación digital de la sociedad, buscando la automatización como rama de la inteligencia artificial, importante para crear sistemas automáticos de forma similar como lo realizaría el ser humano (Tokio School, 2022).

Los conceptos de machine Learning funcionan por medio de un sistema de aprendizaje de tres características, donde el uso de un algoritmo digital es fundamental para emular el aprendizaje. Para ello, se aplica una estrategia de tres procesos conocidos como proceso de decisión, proceso de función de error y el proceso de optimización. En el caso del proceso de decisión se busca establecer el uso de una base de datos de entrada mediante un algoritmo el cual permite analizar el patrón de los datos y por medio de estas datos establecer una predicción. El segundo factor es la función de error, la cual permite desarrollar evaluaciones de predicción del sistema para realizar una comparativa sobre los datos para identificar el grado de predicción respectivo. Mientras que, el proceso de optimización fomenta el uso de los ajustes para el análisis de los datos, logrando reducir las posibles discrepancias entre las predicciones realizadas y el ejemplo conocido (Maisueche, 2019).

Entre las empresas que lideran el desarrollo y uso de inteligencia artificial y Machine Learning se encuentran la IBM, SAS, Datarobot, Microsoft y Google. Estas

empresas han logrado obtener un importante reconocimiento en el mercado gracias a que desarrollan una estrategia de mejorar la eficiencia operativa, se fomenta el aumento de la innovación, se mejora la experiencia del usuario. A pesar de que puedan presentarse ciertos problemas relacionados como son los altos costos, la falta de automatización, la falta de experiencia y la falta de escalabilidad, los cuales podrían generar problemas en la segmentación de la tecnología. No obstante, los sistemas de Machine Learning que existen en la actualidad permiten desarrollar importantes estrategias como el servicio, el desarrollo del registro, el seguimiento y el sistema de monitoreo, los cuales son fundamentales para aumentar el rendimiento óptimo del análisis de datos desarrollado por Machine Learning (Medina, 2023).

El uso del Machine Learning para la inteligencia artificial ha sido fundamental siendo necesaria para analizar la información de los clientes y solucionar problemas relacionados con la producción del comportamiento de los consumidores. Es decir, que los conceptos de Machine Learning han sido desarrollados para el reconocimiento de los patrones, y mejorar los parámetros de los modelos de uso por medio del reconocimiento de los datos previos con los cuales fue codificada para analizar la información. Es decir, que mediante la estrategia de análisis de información basado un ejemplo, los algoritmos serán capaces de analizar todos los datos que son cambiantes y que se pueden encontrar entre ellos para predecir un determinado comportamiento.

El uso del Machine Learning permite desarrollar hasta nueve tipos de tareas distintas que son la clasificación, la regresión, la identificación de las similitudes, el proceso de agrupación, la asociación por coincidencia, la caracterización del comportamiento típico, la predicción de los vínculos, la reducción de los datos y el uso de los modelos causales. Como principales beneficios del uso del Machine Learning en las empresas donde se estudia el comportamiento de los clientes se encuentran el lograr mayor eficiencia. Es decir, que, gracias a las herramientas digitales, el concepto de machine Learning es capaz de emplear con un menor tiempo el análisis de los procesos y el desarrollo de actividades con una mayor rapidez y eficiencia para que las personas dentro de la empresa puedan invertir su tiempo de otra forma, que se más productiva y rentable para los usuarios.

Además, otro de los beneficios que se genera es el aprendizaje constante que se fomenta mediante el estudio de los procesos que en periodos anteriores se hubieran

ejecutado de forma manual. Pero que, gracias a las herramientas digitales estas se desarrollan de forma automática, sin necesidad que realmente una persona física deba comprometerse en guiar o controlar a la herramienta digital. Incluso, las machines Learning actuales son capaces de mantener una conversación con los clientes de forma didáctica gracias al Chat Box, reduciendo el uso de personal.

En el libro de Larsen y Becker (2021), se determina que el concepto de Machine Learning, es necesario que la empresa desarrolle sus propios criterios de evaluación sobre los resultados, para determinar la eficiencia de los proyectos empresariales, siendo que este proyecto va orientado a obtener no un aumento de clientes, en su defecto, se buscará alcanzar a los clientes más adecuados que si estén dispuestos a cumplir con el proceso de aprendizaje de idiomas y experiencia laboral internacional, para garantizar la rentabilidad en la empresa. Entre las preguntas que plantea la estrategia se encuentran sobre que personas o clientes utilizarán el modelo, si la empresa estará de acuerdo con el proyecto, si se pueden utilizar modelos pilotos y su relación costo-beneficio para la empresa (Larsen & Becker, 2021)

Para el sector empresarial, el uso de Machine Learning otorga ciertos beneficios de rentabilidad, puesto que gracias al análisis de las principales características del cliente se puede otorgarle un mejor servicio, permitiendo analizar las preferencias de los clientes, a fin de que se pueda ofrecer productos personalizados de manera automática y sobre una mejor estrategia de percepción para beneficio de la empresa, mejorando su potencia en la fidelización. Con la información adecuada recopilada por la empresa se obtienen datos certeros que disminuyan los errores de la empresa mediante el aprendizaje automático. A medida que esté funcionando por mayor tiempo, mejor será su rendimiento. Machine Learning también ofrece estrategias para las acciones preventivas en el uso de herramientas de aprendizaje, evitando tomar riesgos innecesarios. Incluso, ofrece beneficios en contra de ataques cibernéticos, detección de fraudes y la automatización de los procesos, las cuales aportan con un mejor rendimiento gracias al uso de la inteligencia artificial (Zemsania, 2023).

La información es el principal beneficio en el uso de los conceptos de machine Learning. Puesto que, gracias a los conceptos desarrollados por la teoría de Machine Learning, la empresa EDEX podrá realizar una mejor estrategia para clasificar a sus clientes, basado su perfil de consumo. Además, la información del cliente será

fundamental para analizar los posibles patrones de comportamiento para estudiar y determinar el target más preciso y específico que otorgaría beneficios a la empresa de forma más sostenida. En lugar que la empresa deba invertir en campañas publicitarias y promociones para un público muy general, los conceptos del Machine Learning permitiría analizar el target más específico, con el cual se generaría una mayor probabilidad de generar el proceso de compra. Además, ante este beneficio también se logrará segmentar de forma más eficiente la información y analizar los datos. Todos estos datos podrán ser analizados desde varios lugares como los compradores, bases de datos en internet, procesos de fabricación y todos estos elementos serán analizados para que la empresa basada el análisis realizado pueda mejorar su toma de decisiones con mejor rendimiento (Altertecnica, 2023).

EDEX Education Group al ser una empresa que se ha desarrollado en el entorno de los viajes y de la educación de idiomas de manera nativa, necesita implementar una estrategia en la que se considere la implementación de análisis del mercado y el estudio de sus clientes mediante herramientas tecnológicas. Es importante desarrollar una estrategia donde la empresa EDEX Education Group pueda mantener a todos sus clientes fidelizados, para reducir así la deserción, que causa la pérdida de los ingresos para la empresa.

El uso de los conceptos tecnológicos, siendo de gran importancia como el Machine Learning y la inteligencia artificial como elementos fundamentales en el estudio de los consumidores sobre sus características principales, comportamiento en el mercado y el estudio de factores que son los responsables de generar las variaciones en el comportamiento del consumidor. Machine Learning ofrece cualidades donde la Inteligencia Artificial es capaz de analizar el comportamiento del consumidor y determinar su comportamiento. Los datos que se obtienen permitirán desarrollar nuevas estrategias enfocadas en la atención al cliente que estén acorde a sus reales necesidades reales que puede que se mantengan por un largo periodo de tiempo o que se modifiquen en un corto plazo. El concepto Machine Learning permitirá descifrar de forma eficiente el comportamiento del consumidor.

La Machine Learning es capaz de analizar a todos los consumidores, incluso se puede considerar que podría simular el análisis de los sentimientos de los consumidores. Esto se debe gracias a que la herramienta posee elementos que permite medir la

experiencia del uso y el sentimiento del cliente mediante estos dos elementos que determinan el comportamiento de la experiencia del consumidor. Para ello, se puede depender de los sistemas de back-end, los cuales permiten visualizar y analizar las respuestas de los clientes por medio del texto ingresado, sus criterios de búsqueda y utilización de las fuentes tecnológicas. Con ello, se puede realizar una interpretación del lenguaje natural para clasificar de forma automática los sentimientos de los consumidores, haciendo que la información con la cual se alimenta el machine Learning sea mayor y más precisa para estudiar.

Para el Machine Learning, la opinión de los clientes es un elemento fundamental que garantiza el crecimiento de los negocios para aumentar el rendimiento empresarial. Gracias a estos elementos, las empresas pueden ser reconocidas por ser confiables para los clientes, principalmente porque las empresas han logrado descifrar de forma más efectiva el comportamiento de los consumidores. Mediante la información obtenida por medio del feedback, se puede establecer información cualitativa y cuantitativa sobre la experiencia obtenida al momento de comprar un servicio o utilizar un producto, generando en el consumidor un valor para cada marca (Vásquez, 2019).

El concepto machine Learning para EDEX Education, representaría una oportunidad para segmentar el mercado y ser más preciso para determinar sus necesidades reales. Es decir, que gracias a la herramienta se podrá analizar los datos de los clientes y perfilar de manera automática para dar un puntaje su experiencia y determinar porque han dejado de ser clientes de la empresa. Estudiar la información obtenida permitirá determinar las estrategias y funcionamiento erróneo que estaba realizando EDEX para sus clientes y porque la empresa comience a tener un valor menor para los consumidores. Toda la información recopilada debería ser obtenida por medio del valor de vida medio, los costos de adquisición generados, la satisfacción con el cliente, la tasa de retención que esta pueda generar, el beneficio neto y el proceso de la información para determinar el comportamiento del consumidor.

Para ejecutar estas estrategias se puede implementar el uso de clústeres de K-Means para agrupar la información dentro del algoritmo y establecer una supervisión de la información, a fin de asignar a cada punto un determinado valor y agrupar en torno a las características del cliente. También, se podrá añadir información sobre los hiper parámetros óptimos para determinar la elección en los segmentos de los clientes y

encontrar el grupo de clientes más idóneos que no desertarán del servicio. Con la información recopilada será más eficiente el análisis de los clientes para determinar la rentabilidad y desarrollar las estrategias de campañas de publicidad que vayan directamente al público objetivo que la empresa necesita. Además, a los clientes ya enganchados, se creará una estrategia para mantenerlos como clientes fidelizados y mantener a la empresa en óptimo rendimiento comercial (Brita Inteligencia Artificial, 2023).

En pocas palabras, la inteligencia artificial y el Machine Learning estuvieran siempre vigentes a beneficio de todos los clientes de la empresa EDEX Education Group, como un factor de cambio y desarrollo empresarial donde el rendimiento y el uso de herramientas tecnológicas es fundamental para mantener la competitividad, el crecimiento y las ganancias de una empresa. Además de, ser fundamentales para generar el crecimiento del negocio en el mercado y ser más rentable económicamente para el mercado de idiomas y de turismo, al cual apunta EDEX Education Group.

4. Alcance

La implementación del Machine Learning es esencial actualmente para optimizar los procesos, mejorar la eficiencia y la competitividad de las organizaciones. El público objetivo de este proyecto de investigación son los directivos de la empresa EDEX Education Group para que tomen la decisión de implementar este modelo clasificador para poder tomar decisiones más exactas y obtener resultados eficientes.

La aplicación de esta metodología en la empresa mejorará el servicio al cliente ya que gracias a su efectividad se permitirá analizar las preferencias de cada uno y así se logra potencializar la fidelización. Las personas encargadas de implementar esta inteligencia deben entender que gracias al Machine Learning se pueden automatizar los procesos lo que conlleva a que se cierren ventas en menos tiempo del que ellos están acostumbrados.

Por otra parte, con el uso diario de esta nueva herramienta se pueden realizar acciones para prevenir errores y no tomar decisiones que puedan afectar o al cliente en sí o a toda la organización. Como EDEX es una empresa que se dedica a la venta de cursos en el extranjero, el Machine Learning puede ser de mucha ayuda para detectar transacciones ilegítimas de dinero.

La presente investigación es una necesidad para las personas o empresas que quieren adquirir conocimientos sobre cómo aplicar correctamente las diferentes herramientas que brinda la Inteligencia Artificial para que logren interpretar los datos y agruparlos de la manera correcta con la finalidad de poder detectar los errores y descubrir cómo solucionar o mejorar dichos procesos.

Gracias al Machine Learning se puede hacer retrospectiva en los errores del pasado para no volverlos a cometer. La mayoría de las empresas en Ecuador no implementan el Machine Learning, debido que su confianza en la tecnología es limitada, alcanzando solo el 12% de las empresas en el mundo que la utilizan como herramienta para superar a los competidores. La inversión tecnológica ante el uso de la tecnología es limitada y en el Ecuador, solo se aprovecha por las grandes corporaciones como Corporación Favorita, Nestlé, Pronaca, Cervecería Nacional, Toyota, La Fabril y el sector Bancario con empresas como Banco Pichincha, Banco Bolivariano, Banco de Guayaquil y Produbanco, como una estrategia de análisis para determinar la rentabilidad de sus clientes y las posibles variaciones en el mercado (ASOMIF, 2022).

Se estima que más del 40% de las organizaciones en los Estados Unidos de América utilizan esta tecnología para obtener mejores resultados, tomar decisiones más exactas y facilitar el análisis de sus datos (Morgan, 2019). Son muchas las ventajas de esta inteligencia y los beneficios que se obtienen con su implementación, pero hay que tomar en cuenta el presupuesto de la empresa y las capacidades de sus trabajadores porque es una tecnología con un precio que puede llegar a ser difícil de costear y difícil de comprender. Una de sus herramientas más utilizadas en la actualidad es el Chatgpt, el cual es una herramienta que fundamenta mejorar la capacidad productiva de los trabajadores con el uso de las herramientas tecnológicas, siendo que la productividad empresarial es importante como estrategia de crecimiento para los economistas en el mercado, como una forma de superar los obstáculos en el mercado, aumentando su productividad y generar eficiencia en el desarrollo empresarial. Ante el uso de la herramienta de Chatgpt, en empresas que la utilizan como herramienta de gestión para realizar segundas tareas, las empresas notan que los trabajadores dentro de las empresas realizan de mejor manera sus actividades laborales, aumentando la rentabilidad empresarial (Winn, 2023)

5. Objetivos

5.1.Objetivo General

- Crear un modelo clasificador para oportunidades de captación en el exterior por Machine Learning para la empresa EDEX EDUCATION GROUP

5.2.Objetivos Específicos

- Comprender los conceptos del Machine Learning por medio del marco teórico
- Explicar detalladamente y entender la metodología del Machine Learning
- Analizar la base de datos de la empresa EDEX Education Group del año 2021

6. Preguntas de investigación planteadas

1. ¿Cuáles son los beneficios de crear un modelo clasificador mediante Machine Learning para la empresa EDEX?
2. ¿Cuáles son los problemas específicos en la base de datos presentada por la empresa?
3. ¿Cuál es el impacto de la creación de un modelo clasificador en la mejora del rendimiento de la empresa?
4. ¿Cómo lograr que los asesores se concentren en un grupo específico de estudiantes?

II. CAPÍTULO I: MARCO TEÓRICO

1. Globalización

El concepto de globalización ha sido desarrollado como una terminología que describe crecimiento y cohesión de los estados en el mundo, causando un impacto en el choque cultural y la cohesión del desarrollo económico en el futuro de la economía mundial. La globalización se trata de un evento histórico que ha sido generado como resultado de la innovación del ser humano y el progreso tecnológico. Lo que ha causado una creciente integración sobre las economías del mundo, el desarrollo económico, los flujos financieros y la cohesión de la población gracias al desplazamiento de las personas para buscar nuevas oportunidades laboral o simplemente entretenerse mediante los viajes entre amigos, familiares o de forma individual para aprender sobre los aspectos culturales, políticos y ambientales.

Cabe destacar que la globalización no se trata de un fenómeno que recién comenzó con la llegada del internet, que es como se suele pensar. En su lugar, la globalización realmente comenzó desde el proceso de industrialización de la humanidad, hace más de cien años, cuando los países vieron necesario la búsqueda de trabajadores para desarrollar los mercados industrializados y agrícolas, siendo la migración desde Europa hacia América como el inicio de la cohesión entre distintas culturas y costumbres en el mundo. A inicios del siglo XX la cohesión entre las distintas personas solo se podía realizar de manera personal. Sin embargo, con la llegada del internet, los servicios y sistemas financieros han comenzado a generar una mayor cohesión e integración, permitiendo que los mercados financieros sean cada vez más dependientes de los demás. Este hecho ha causado que exista un crecimiento económico jamás visto, incrementando el PIB de las naciones en el mundo. No obstante, este crecimiento no se realizó de forma sostenida ni regulada, causando que la mayor expansión y cohesión de la humanidad sea más notoria en la mitad del siglo XX con el desarrollo de las industrias petroleras y a finales del siglo XX con la creación del internet, responsables de impulsar nuevos modelos de comunicación y comercio (Fondo Monetario Internacional, 2020)

La globalización representa la generación de nuevas oportunidades de crecimiento en el mundo, garantizando el acercamiento y la conexión de los mercados, las sociedades y las culturas en el mundo. Sin embargo, esta misma globalización y sistema de

interconexión entre los estados es el principal factor que permitió que enfermedades como el COVID 19 afectaran a millones de personas en todo el mundo, gracias a los sistemas de transporte que fomentan el intercambio de información y personas de manera constante. Gracias al fenómeno de la globalización, se considera que el mundo como un único entorno en el cual cohabita la población en general en un mismo ambiente, permitiendo que la tasa de pobreza mundial pueda disminuirse hasta en un 84% desde 1980 (Raya, 2021)

2. Mercado Laboral

El mercado laboral es uno de los elementos más importantes de la economía, puesto que representa la unión de la oferta y de la demanda en la generación de empleo en un lugar determinado. Siendo un factor referente para establecer el punto de encuentro entre el empleo disponible y las personas que buscan el empleo considerado como ideal para fomentar un mejor estilo de vida. El mercado laboral también representa la necesidad del sector empresarial e industrial para que las empresas basadas en sus necesidades comiencen con la búsqueda de personas con habilidades indispensables para el desarrollo empresarial que fomenten su crecimiento en el mediano y largo plazo. El desarrollo del mercado laboral suele funcionar dependiendo de las necesidades existentes en el mercado sobre determinados productos y servicios. A fin de generar el interés de una empresa sobre la dependencia de trabajadores basados sus capacidades e interés de contratación. Una vez que la empresa encuentra a su trabajador ideal, se deberá establecer un proceso de negociación para garantizar un salario digno y el tiempo que invertirá el trabajador para el desarrollo de la empresa (Arranz, 2022).

El mercado laboral al igual que el mercado suele cambiar constantemente basadas las tendencias tecnológicas y las necesidades del mercado existentes. El mercado laboral se encuentra en constante evolución, generando que las empresas deban evolucionar, transformarse y adaptarse a las nuevas necesidades de la sociedad. Uno de los más importantes requerimientos que se centran en el mercado laboral actual son las competencias digitales para el nuevo mercado laboral, siendo necesario que los trabajadores tengan capacidad para desarrollar sus habilidades cognitivas según el entorno tecnológico, el desarrollo del aprendizaje sobre los programas y generar aplicaciones que causen un valor de mercado positivo. No obstante, uno de los cambios más notorios existentes en el mercado laboral actual se centra en la necesidad de la

flexibilidad y la conciliación dentro del entorno de trabajo. La flexibilidad se ha convertido en un acto fundamental para los lugares de trabajo debido que estos se caracterizan por contar con horarios flexibles, teletrabajo y una relación directa entre los altos mandos y los trabajadores gracias a una comunicación horizontal. Ante esta necesidad, las empresas han desarrollado una estrategia donde la flexibilidad es fundamental para garantizar que el trabajo se realizará de forma eficiente, incluso combinando acciones como el teletrabajo y el trabajo presencial aplicándose sistemas de rotación. El trabajo en equipo es otro de los factores principales que valora el sector empresarial, gracias a que existe una necesidad de fundamentar un ambiente de trabajo eficiente con una correcta cohesión entre los trabajadores y la empresa, sin importar que se trate de un sistema autónomo o una gran corporación, creando espacios de trabajo en equipo que esperan fundamente el crecimiento efectivo del negocio (Pop, 2023).

3. La importancia de los idiomas

Uno de los aspectos más importantes del mercado laboral actual son los idiomas. Que un trabajador sea capaz de manejar al menos un idioma adicional de su lenguaje autóctono, le fomenta una mayor competitividad, un mejor acceso a los cargos públicos y una mayor capacidad para ser promovido dentro del trabajo a menores plazos, causando una mejor capacidad para obtener mejores puestos de trabajo y mejores ingresos que los demás. El conocer un idioma adicional genera en los trabajadores una mayor competitividad gracias a que los postulantes tendrán una mejor capacidad sobre distintos puestos de trabajo, y el idioma es tan importante como que el trabajador haya realizado distintos estudios de postgrado o posea una experiencia en el mercado laboral del exterior.

El idioma más importante y utilizado en el mercado laboral es el inglés, puesto que se ha convertido gracias a la influencia de Estados Unidos en el mundo, en el idioma de comunicación internacional, de comercio y de finanzas de mayor uso en el mundo. El inglés es considerado como la lengua franca dentro de las Organizaciones Internacionales como son la Unión Europea, la Unesco o las Naciones Unidas. Por ende, aprender el idioma es fundamental para garantizar el éxito en el mercado laboral. Más aún si se comienza con el aprendizaje de un idioma adicional al inglés, sea el francés, el mandarín, el alemán, entre otros. La necesidad de aprender nuevos idiomas en las empresas se ha convertido en una necesidad causada por la globalización de la economía y de la búsqueda continua en el compartimento de nuevas prácticas entre sucursales que existen sobre el

dominio de otro idioma. Cuando un trabajador no conoce el idioma, su conocimiento se ve limitado debido que no se puede generar una interacción correcta entre empresas, trabajadores o clientes. Las empresas necesitan de forma importante trabajadores que sepan nuevos idiomas y es por ello que, en la actualidad, 3 de cada 4 puestos laborales establecen como requisito el manejo del idioma inglés para mantenerse competitivos en el mercado laboral (Seminarium Certificación, 2023)

4. Importancia de la formación académica de alta calidad

Una formación académica de calidad representa una oportunidad para que los futuros trabajadores obtengan un crecimiento de intelectualidad que les garantice desarrollar sus capacidades analíticas, críticas y preparase mejor para la resolución de los problemas. Una formación académica de calidad genera en las empresas un inminente interés para las empresas en contratar siempre a los trabajadores académicamente mejor preparados. La formación profesional representa en las características que puede tener una persona para fomentar su inserción laboral tanto en el corto y mediano plazo. Una vez que la persona que forme parte de una plantilla laboral comenzará su periodo de preparación combinando la formación teórica con la práctica, fomentando así su desarrollo y experiencia laboral (Carlos de la Hoz Academia, 2019).

Dentro de la formación académica, es la formación universitaria conocida como el grado educativo más importantes, debido que representa el desarrollo de la investigación y el conocimiento de las personas, enfocándose en su desarrollo profesional dentro del mercado laboral. Las formaciones academias superiores otorgan una mayor oportunidad de superación, para obtener puestos de trabajo de mayor rango dentro del mercado laboral. Para aumentar el conocimiento cognitivo, las personas pueden desarrollar aún más sus capacidades por medio del desarrollo de postgrados también conocidos como masterados y doctorados. En el caso de los estudios postgrado, representan el acceso a un mayor grado de especialización dentro del sector académico, el cual representa que la persona se ha formado con mayor conocimiento y capacidad de desenvolvimiento dentro del mundo laboral. Esta preparación académica suele tardar un promedio de dos años de preparación. Mientras tanto, el doctorado representa una mayor formación académica gracias a que las personas consiguen desarrollar su tesis doctoral por medio de la práctica y desarrollo de sus trabajos de investigación, logrando la mayor

distinción posible en el entorno académico. Los doctorados son el título de mayor rango dentro de los mencionados (Mundana, 2016).

La formación académica para el entorno laboral es importante debido que representa una mejor calidad de vida gracias a que se puede contar con una mayor capacidad de desarrollo formativo y obtención de un puesto de trabajo de alto reconocimiento. Sin embargo, cabe destacar que la formación académica en algunos casos, principalmente en países de América Latina, la preparación académica no es el factor más importante para la obtención de un título. En su lugar, se considera de mayor relevancia la experiencia laboral e incluso las relaciones que tenga la persona con alguna otra que le permita obtener un determinado puesto de trabajo. Solo por establecer un ejemplo, en el Ecuador, el 33.6% de sus estudiantes graduados y egresados no logra conseguir un puesto de trabajo digno (Zapata, 2021). Este problema no solo ha afectado a América Latina, puesto que en España también se evidencia este problema con una incidencia de al menos el 30% de los universitarios graduados, los cuales se les dificulta en gran manera conseguir un puesto de trabajo luego de 4 años de haberse graduado. Siendo así que, la importancia de tener un título es un factor menos relevante o que represente una garantía para obtener un puesto de trabajo adecuado (Diario El País, 2019).

5. Importancia de la experiencia laboral

La experiencia laboral es un elemento altamente valorado por las empresas dentro del mercado laboral. La experiencia laboral representa en la oportunidad para los trabajadores de adquirir habilidades y conocimientos que no se pueden aprender dentro del sistema educativo. Además, la experiencia laboral representa la oportunidad de poner en práctica los conocimientos adquiridos en un entorno de trabajo real. Es la experiencia laboral el factor que genera beneficios en los trabajadores que van desde la exposición de los hábitos de práctica profesional, el desarrollo de una mayor conciencia sobre sí mismo, el desarrollo de las oportunidades para desarrollar la responsabilidad cívica y otros aspectos que fomentan el desarrollo profesional y aumentan el curriculum de los trabajadores (Westreicher, 2020).

Existen ciertas habilidades que forman parte de la experiencia laboral de las personas, las cuales son el manejo del estrés, el autoconocimiento, la empatía, la comunicación, el desarrollo del pensamiento crítico, desarrollar el asertividad, la

creatividad, el desarrollo de las relaciones interpersonales y el manejo de los problemas en la toma de decisiones. Estas habilidades son fundamentales de ser explotadas para que los trabajadores puedan gozar de una mejor calidad en su trabajo. Solo por establecer las características, en el mercado laboral de México, solo el 26% de todos los trabajadores obtiene su primer puesto de trabajo aun estudiando la universidad. Mientras tanto, el 38% lo consigue en menos de seis meses. En el mercado laboral mientras más tiempo una persona tarda en conseguir empleo, será más difícil desarrollar su capacidad profesional. Además, la experiencia es un requisito fundamental en las empresas de América Latina, siendo este el factor que determina las variaciones de los ingresos o generar menores oportunidades para obtener un puesto de trabajo adecuado (Editorial RSyS, 2022).

Solo en América Latina, existe un problema relacionado entre la tasa de desempleo de los adultos jóvenes entre 15 a 24 años de edad y la media de los adultos de entre 25 a 65 años. Los cuales consideran que la tasa de empleo es más alta para los jóvenes mayores de 25 años con una tasa de empleo en Uruguay que alcanza el 65.4%. Mientras que los menores de 25 años de edad alcanzan una tasa de 35.5%. Para el sector empresarial de América Latina, la experiencia laboral es un factor fundamental, incluso siendo superior que la calidad educativa. Sin embargo, a medida que los trabajadores alcanzan una mayor edad, su capacidad para obtener un puesto de trabajo digno se ve afectada porque tienen menores oportunidades de lograr obtener un empleo. En el caso de México, esta cifra representa que la tasa de desempleados que poseen estudios de educación superior alcanza el 5.46%, comparado con el 2.84% de las personas desempleadas con la primaria sin completar. Es decir, que se debería tener mayores oportunidades laborales con una mayor formación académica. Sin embargo, para las empresas el factor más importante es la experiencia laboral. Siendo fundamental en el sector que, ante una posible falta de experiencia previa en las carreras, se generen dificultades para que las personas recién graduadas o con un proceso académico en formación cuenten con menos capacidad para acceder a un empleo digno (Golpe, 2015).

6. Empresa EDEX EDUCATION GROUP

La empresa EDEX Education Group fue fundada el 29 de enero del 2021 con su sede principal en la ciudad de Quito. La empresa se encuentra ubicada en el sector del mercado de idiomas, convirtiéndose en una empresa consultora de estudios en el extranjero, fomentando el trabajo en equipo para que la sociedad ecuatoriana pueda

alcanzar su sueño de estudiar idiomas en el extranjero o el acceder a un programa universitario por medio del asesoramiento personalizado que EDEX ofrece a su público objetivo.

EDEX Education Group ofrece estrategias de asesoramiento personalizada para el estudio del idioma inglés y acceder a Universidades y Academias con altos estándares educativos de educación superior en el extranjero. EDEX ha sido formada por consultores académicos profesionales que poseen una amplia experiencia internacional sobre cada destino otorgado a sus clientes. El trabajo de la empresa se enfoca en buscar todas las facilidades posibles para que los estudiantes puedan acceder a una educación de alto nivel, logrando expandir sus oportunidades laborales hacia el mercado internacional. La empresa posee una alta experiencia en el mercado de trabajo, permitiendo que sean los jóvenes los que se han beneficiado de vivir en el extranjero para obtener de forma directa la experiencia de vivir, estudiar y trabajar fuera del territorio nacional durante algunos meses o años, según sea el requerimiento de los clientes y la elección del destino (EDEX Education Group , 2023).

EDEX Education Group es una marca que se encuentra totalmente comprometida con la satisfacción de las metas y los sueños de los estudiantes por medio de las Instituciones de calidad, buscando que los estudiantes logren alcanzar sus metas y desarrollo de los miembros del equipo de trabajo. Esto se logra por medio de la interacción entre las instituciones académicas de alta calidad con las cuales EDEX Education Group cuenta con alineaciones directas para un óptimo desarrollo. La empresa también se compromete con fomentar la acción de responsabilidad, aplicándola sobre las normativas existentes en todas las embajadas con las cuales se maneje un destino de estudio, al igual que con las diferentes instituciones para que el estudiante pueda cubrir todas sus necesidades y expectativas (EDEX Education Group , 2023).

EDEX Education Group se ha convertido en una empresa que siempre plantea su estrategia de acompañamiento a los clientes, para fomentarles siempre un desarrollo eficiente del progreso académico y que se logre de la mano con la formación de las instituciones con las que el estudiante pueda mantenerse seguro y que sea recibido de forma positiva para una óptima formación académica. La calidad con la cual se forma los estudiantes es fundamental debido que representa el reconocimiento académico internacional para la empresa, fomentando los estudios en el extranjero de las personas

que siempre planteen como meta en el mercado laboral el gozar de un mayor desarrollo cognitivo, crecimiento en el área de idiomas y experiencia laboral.

Como estrategia de valor de la empresa, EDEX revisa el perfil de los clientes, determinando sobre sus gustos, sus intereses, sus metas y sus estudios, con el objetivo de lograr una mejor prestación y propuesta académica a la vanguardia del mercado. EDEX también ofrece otras ventajas respecto al mercado educativo puesto que permite la creación de proyectos por medio de la definición de los programas, instituciones de destino y todo el equipo de trabajo con el cual se laborará para completar el proyecto de estudios en el extranjero de forma adecuada para beneficio de los clientes (EDEX Education Group , 2023).

Es decir, EDEX Education Group no solo se enfoca en el conocimiento del mundo de los idiomas. Puesto que, también ofrece beneficios a los clientes relacionados con los cursos de pregrado, diplomado, certificaciones, maestrías e intercambio educativo, siendo la organización una marca que otorga asesoría a sus clientes. La asesoría otorgada es gratuita para todas las personas interesadas en el servicio. Actualmente, EDEX Education Group posee más de 8 mil estudiantes repartidos en más de 300 instituciones en el mundo. La empresa goza de 12 años de experiencia en el mercado (EDEX Education Group , 2023).

7. Inteligencia Artificial

La inteligencia artificial consiste en el desarrollo y la combinación de algoritmos informáticos, los cuales tienen como principal función la de desarrollar una función específica basada las necesidades de su creador. El funcionamiento de la inteligencia artificial se logra por medio del uso de sistemas informáticos que, trabajando entre sí, logran imitar la función cognitiva del ser humano. Para ello, se utilizan maquinas, procesadores y softwares para realizar determinadas tareas de procesamiento y análisis de la información. La inteligencia artificial se trata de la codificación del ser humano para desarrollar el pensamiento del ser humano utilizando los sistemas de cómputo, permitiendo al organismo digital a razonar, aprender y realizar acciones por su cuenta, a fin de que pueda resolver problemas de forma directa (Ferrovial, 2023).

La inteligencia artificial se clasifica según su funcionamiento de los cuales se encuentran la inteligencia artificial reactiva, la inteligencia de memoria limitada, la inteligencia artificial de teoría de la mente y la inteligencia artificial de la de autoconciencia.

La inteligencia artificial reactiva se considera como el tipo de inteligencia artificial más sencillo que existe debido a su desarrollo. Puesto que, el sistema ha sido desarrollado con la finalidad de generar decisiones sobre un análisis en tiempo presente, sin necesidad que esta tenga una memoria para captar el conocimiento. Es decir, este sistema solo tomará una decisión en base a un evento presente, evitando analizar datos pasados de su experiencia que mejoren su capacidad de desarrollar un resultado más acertado. Siendo así que estos sistemas tienen poca capacidad de evolucionar debido a su sistema que no genera el crecimiento de su capacidad cognitiva. Uno de los sistemas más utilizados de la inteligencia artificial reactiva es el denominado como Deep Blue, siendo este un sistema que, a pesar de haber sido capaz de ganarle al campeón de Ajedrez Garry Kasparov, es un sistema que solo logró la victoria procesando la información en tiempo real basado los movimientos de su oponente (Grupo Atico34, 2023).

La inteligencia de memoria limitada se caracteriza por ser un tipo de desarrollo de inteligencia artificial orientado a obtener datos pasados, pero de una manera limitada para analizar la información y tomar con ello una correcta decisión. Es decir, que el sistema de inteligencia artificial de memoria limitada intentará simular la recordación como una forma de comportamiento que mejorará su sistema de análisis situacional. Permitiendo ante ello mejorar su comportamiento y anticiparse ante posibles tareas, funciones y requerimientos que el usuario necesite. Este sistema imita un comportamiento de recordación que garantiza que la toma de decisión aplicada sería más eficiente en comparación que la inteligencia artificial reactiva. Sin embargo, esta memoria es limitada, debido que su sistema no ofrece una mayor complejidad en el razonamiento, ni tampoco ofrece un mayor esquema de pensamiento humano. Son los algoritmos diseñados en el sistema de memoria limitada, como un sistema de recordación básico que mejoran las tareas, pero que no pueden garantizar que el sistema será altamente eficiente, adaptándose únicamente a los servicios más simples como sería el uso de vehículos inteligentes que laboren en el mercado de la mensajería (DispatchTrack, 2019).

La inteligencia artificial de teoría de la mente se trata de un sistema de inteligencia artificial que simula realmente todas las implicaciones sobre el pensamiento del ser humano para ser replicados mediante sistemas informáticos. La teoría menciona que el desarrollo de la inteligencia artificial de teoría de la mente está basado en un entorno no comprobado, puesto que solo se ha generado en hipótesis. Sin embargo, se establece que el desarrollo de las máquinas que sean capaces de comprender las emociones, deseos y las necesidades del ser humano, serán consideradas como sistemas informáticos de teoría de la mente, replicando casi a la perfección el pensamiento del ser humano y la interacción social (UTECH, 2018).

La inteligencia artificial de la de autoconciencia se considera el máximo exponente del desarrollo de la inteligencia artificial. Esta se refiere a la capacidad de desarrollar un sistema en el cual se logra mantener la comprensión del sistema sobre sí mismo, sobre su entorno y toda la composición que existe en él, logrando adaptarse a sus necesidades y aprender por medio de la interacción según lo requiera. Este tipo de inteligencia artificial aún se mantienen lejos de desarrollarse, principalmente a la complejidad que se demuestra en el pensamiento humano. Uno de los ejemplos cinematográficos más referentes sobre la autoconciencia de los sistemas informáticos es el caso de Skynet, el cual mediante su propio análisis concluyó que son los seres humanos el mal que afecta al planeta, viendo necesario su erradicación (Páramo, 2017).

8. Inteligencia artificial en el entorno económico y mercado

Es la propia inteligencia artificial el factor que ha generado cambios en el entorno de la economía y en cualquier aspecto existente en la vida cotidiana. La razón se debe a que la inteligencia artificial busca realizar todas las labores que el ser humano forma parte, permitiéndole un mayor rendimiento en los resultados, sin importar que realmente pueda afectar al largo plazo su implicación en los puestos de trabajo, en la industria y el mercado en general. Desde que se comenzó a desarrollar sistemas de inteligencia artificial en el sector industrial, el sector ha dependido menos de los seres humanos, reduciendo los puestos de trabajo en el mundo. Actualmente, la inteligencia artificial ha sido utilizada para desarrollar automóviles que no dependen de conductores humanos, lo que generaría la eliminación al mediano y largo plazo del trabajo de taxistas o de choferes de buses.

Los cambios en la inteligencia artificial se encuentran ligados de forma directa con la toma de decisiones que formen parte en el desarrollo de la inteligencia artificial. En la actualidad, los principales sectores que se benefician del desarrollo de la inteligencia artificial son el sector de la salud, transporte, educación, servicios y el sector financiero. Siendo este último utilizado para analizar el comportamiento de las variables del mercado para determinar el rendimiento empresarial futuro esperado en el largo plazo (Colomé & Navajas, 2021)

A pesar de los problemas en el mercado laboral que podría significar la inteligencia artificial, existen ciertos aspectos positivos que son evidentes, como su utilización en el entorno contable y en el manejo de las finanzas de la banca. Puesto que, en estos sectores, se necesitará de un menor tiempo para analizar elementos como el pago de las facturas, la realización de los informes financieros o cualquier otra tarea indispensable en el sector empresarial. La inteligencia artificial es capaz de analizar los datos de la empresa y aumentar su capacidad de automatización, generando en los trabajadores la posibilidad de mejorar en sus nuevos talentos y habilidades.

El mercado de desarrollo de la inteligencia artificial en la actualidad es capaz de generar ganancias por 5.05 mil millones de dólares, en comparación con los 420 millones que generaron en su desarrollo desde el 2014. Su mayor desarrollo en el proceso de aprendizaje de los sistemas se debió a la elaboración de las Machine Learning, convirtiéndose en un elemento indispensable en el sector empresarial debido a que estos sistemas se puedan adaptar a las nuevas necesidades de cada empresa. Además, existe un promedio de 6 mil millones de dispositivos en la actualidad que utilizan Inteligencia Artificial para realización de sus actividades cotidianas u obtener respuesta de interrogantes. Una de las Inteligencias Artificiales más utilizadas en la actualidad es el sistema Chatgpt (Fundación Aquae, 2023).

El sector privado se ha beneficiado de forma importante gracias al desarrollo de la inteligencia artificial. En promedio, existen más de un millón de empresas en el mundo que se dedican a la elaboración de herramientas de inteligencia artificial y el concepto de Machine Learning, enfocadas en las necesidades de cada negocio. Estas empresas realizan inversiones en promedio de 5.4 mil millones en inversión de capital para fomentar el desarrollo de los sistemas. Entre las empresas más importantes se encuentran la empresa Autónoma, enfocada en el negocio de suboptimización mediante el uso de

inteligencia artificial y automatización en el proceso de toma de decisiones, mejoras en las operaciones y organización del servicio empresarial (Pegsystems, 2023).

9. Machine Learning

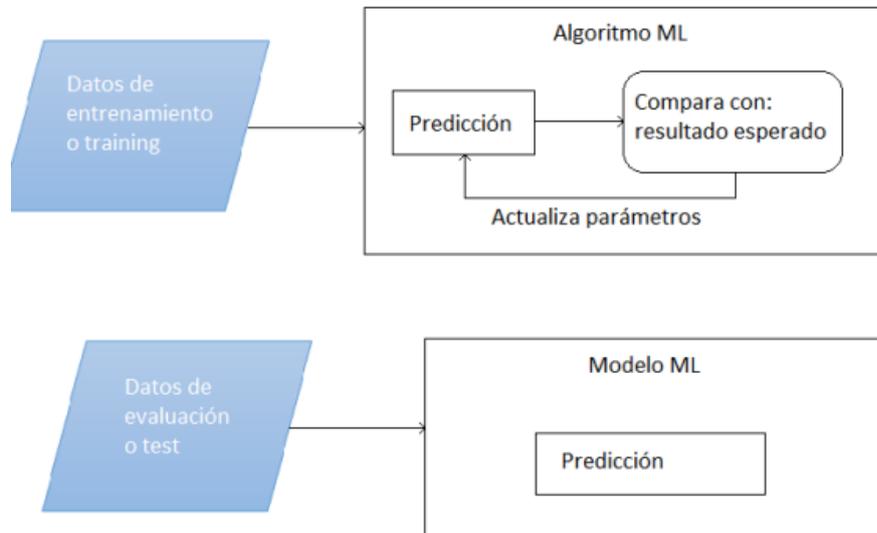
Un concepto de Machine Learning consiste en un sistema tecnológico que está orientado al uso de algoritmos y sistemas estadísticos con los cuales se logra la realización de determinadas tareas en un tiempo de producción altamente eficiente. Su funcionamiento se logra gracias al uso de patrones e inferencias que existen en los sistemas, convirtiendo al sistema de computación con algoritmos en un sistema de Machine Learning que posee y utiliza la información relevante para procesar grandes cantidades de datos para obtener un resultado. Entre la información que se analiza se encuentran datos históricos antiguos y actuales, para que el sistema pueda encontrar una posible inferencia y lograr un resultado preciso, eficiente y obtenido en forma eficaz. Los conceptos de Machine Learning son altamente importantes debido que son elementos necesarios que analizan el crecimiento de la empresa, generando nueva información y resolviendo los problemas complejos en el sistema. Impulsando así el crecimiento empresarial a beneficio del desarrollo de clientes, empleados y desarrollo financiero (Amazon, 2023).

El desarrollo inicial del Machine Learning comenzó gracias al desarrollo de la idea por parte del matemático Walter Pitts y de la idea de Warren McCulloch, en 1943. Su idea consistía en la creación de la inteligencia artificial utilizando la tecnología para simular el comportamiento del cerebro, siendo este comparado con un organismo computacional. Gracias a su idea, en 1950, el científico Alan Mathison diseñó el Test de Turing, siendo un factor importante para determinar el rendimiento de una computadora basado su capacidad de razonamiento e imitando el comportamiento del ser humano. Para el año de 1952, se desarrolló el primer sistema informático considerado con capacidad de aprendizaje.

Este sistema era un software que tenía la capacidad de jugar damas, aprendiendo por medio de un sistema que almacenaba la información sobre los oponentes, estilos de juego y comportamientos del adversario, para según esto mejorar su respuesta y por ende su nivel de juego, mejorando en los resultados a medida que iba jugando. Simulando así el comportamiento del ser humano. En 1956, los informáticos Martin Minsky y John

McCarthy desarrollaron una conferencia científica en la que otorgan el desarrollo de estos sistemas como Inteligencia Artificial. Además, en el mismo año, Frank Rosenblatt fue el diseñador de la primera red neuronal creada mediante las computadoras. Este mismo sistema sería utilizado posteriormente para la creación del internet (Velasquez, 2022).

Figura 1 Funcionamiento de Machine Learning



Nota. Desarrollo de un algoritmo de producción básico del Machine Learning. Recuperado de (Velasquez, 2022)

Las bases para el desarrollo de un concepto de Machine Learning constan de cumplir con directrices como contar con un mismo dominio, el cual debería ser de interés específico de lo que los diseñadores desean alcanzar. El segundo aspecto es el contar con un espacio de hipótesis que conste de todas las posibilidades de soluciones. El tercer aspecto es fomentar el entrenamiento inicial del sistema, a fin de que pueda recolectar datos y aumentar su desarrollo de soluciones distintas a medida que desarrolle su comportamiento. Se deben considerar los criterios de operatividad, elementos que determinan el reconocimiento del tipo de dominio con el que funciona. Los sistemas de Machine Learning solo pueden funcionar con algunas variables ingresadas, pero les permiten almacenarlas para luego desarrollar y formular sus propias variables (Sandoval, 2018).

Para el año 2006, el desarrollo del Machine Learning había alcanzado su cúspide en el manejo y procesamiento de la información, gracias a la intervención de empresas

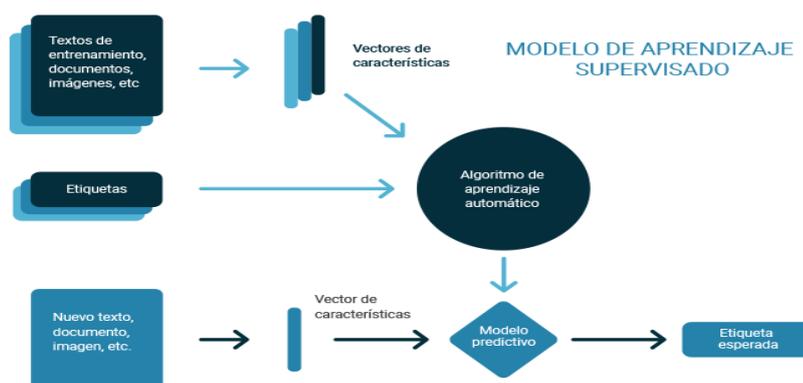
importantes de manejo de datos informáticos como IBM y Microsoft. Permitiendo así que los conceptos de Machine Learning pudieran expandirse su función y utilidades para una mayor cantidad de usuarios, sin importar que se trate de grandes empresas, o empresas Pymes. Una de las mayores asociaciones para el desarrollo de una Machine Learning de alto impacto fue el caso de la asociación de Jeff Dean para trabajar con Google en la creación de una estrategia de innovación denominada como Google Brain, la cual intentaba desarrollar una red neuronal mediante Google, para garantizar la detección de los patrones mediante videos e imágenes (Hinestroza, 2018).

10. Conceptos de Machine Learning

Existen varios tipos de machine Learning los cuales se fundamentan en función de sus implicaciones como son el aprendizaje supervisado, el aprendizaje no supervisado y el aprendizaje con refuerzo.

El aprendizaje supervisado posee etiquetas de conjuntos con los datos para determinar los patrones con los que se pueda detectar la información, al mismo tiempo que se logre utilizar los datos, etiquetando nuevos conjuntos en la información. Es este tipo de proceso de aprendizaje del Machine Learning el cual depende de la utilización de un tipo de algoritmo el cual cuente con distintos patrones de aprendizaje. Sin embargo, se aplica de forma iterativamente sobre los datos para obtener la información según los ordenadores que estén en funcionamiento para obtener los datos escondidos, sin que exista realmente la necesidad de programar el sistema sobre cada uno de los datos obtenidos (Luna, 2018).

Figura 2 Modelo de aprendizaje supervisado



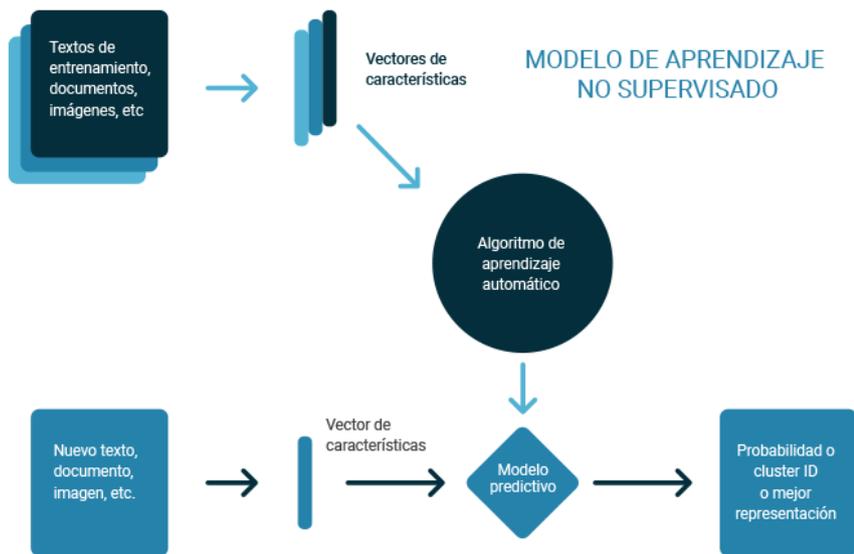
Nota. Ejemplo de un diagrama de flujo de un modelo de aprendizaje supervisado para determinar el comportamiento predictivo de la información. Recuperado de (Luna, 2018)

Para ello, es necesario generar un modelo capaz de predecir el resultado de salida de la información según el ejemplo histórico dado antes. Estos modelos se comportan y se desarrollan a partir de los algoritmos de Machine Learning sobre determinadas características y atributos existentes sobre los datos de entrenamiento. Además, estos sistemas buscan modelar las relaciones de dependencia sobre los resultados y su factor predictivo en base a los objetivos y a las características de entrada generados (Aprende IA, 2023)

El modelo de aprendizaje no supervisado se caracteriza por ser una herramienta informática, siendo la más utilizada de los conceptos del Machine Learning, gracias a su capacidad para aprender y determinar los datos en función a la información obtenida. A diferencia del modelo supervisado que se encarga de etiquetar los datos en conjuntos según sus características, el sistema no supervisado no necesita ningún tipo de categorización o etiqueta de datos para desarrollar su propio proceso de aprendizaje. Para lograrlo utiliza elementos característicos entre los datos obtenidos que cataloga como similitudes. Entre estas se encuentran los valores encontrados, estructuras y patrones dentro de la información. Para el sistema no existirá un dato correcto o incorrecto, determinando solo los patrones según el comportamiento desarrollado con algoritmos amplios, generación de clústeres y asociaciones entre las variables.

Los algoritmos de asociación de clústeres con datos no supervisados presentan una forma eficiente de analizar los datos, agrupando la información mediante diferentes subconjuntos conocidos como clústeres. Se ha convertido en una forma de describir los datos mediante la estructura de información, determinando así las características en los sistemas sobre las mismas características y asignación de los clústeres. Estos clústeres se pueden agrupar de manera jerárquica, por medio de k-medias o utilizando clústeres de difusos o mezcla gaussiana (Luna, 2018).

Figura 3 Algoritmo Machine Learning no supervisado



Nota. Análisis del algoritmo no supervisado para Machine Learning. Recuperado de (Luna, 2018)

Los algoritmos de clúster jerárquico se representan con el uso de Machine Learning que permitan agrupar las cosas con los árboles de clústeres mediante datos agrupados y dividiendo la información sobre grupos más pequeños. Con ello, se logra el resultado final, clasificando los datos genéricos de los específicos y determinando un comportamiento sobre cada grupo (TIBCO, 2023).

El sistema algoritmo de K-medias se caracteriza por ser un sistema donde se separan los datos en distintos clústeres que no se han logrado etiquetar. Ante ello, y según la precisión en la asociación de los datos que se encuentren más o menos cercanos entre sí, se establecerá el nivel de pertenencia sobre el clúster. Una K que sea amplia significará que este forma parte de un grupo más pequeño. En este sistema, cada clúster será asignado con un solo punto de datos (TIBCO, 2023).

En el caso del modelo de mezcla gaussiana, se evidencia una acción de distribución sobre la curva de campana y la de los clústeres, agrupados entre sí sobre las densidades sean estas previstas y siempre mostrando la información sobre los datos generales.

La agrupación de los clústeres de manera difusa se muestra bajo la superposición de por cada punto de datos, aplicarse de forma precisa la información del clúster de manera relevante en contraposición del clúster que no sean difusos sobre los puntos de datos que solo pertenecen a un clúster (TIBCO, 2023).

Figura 4 Algoritmo Machine Learning por refuerzo



Nota. Análisis del algoritmo Machine Learning por refuerzo. Recuperado de (Luna, 2018)

El modelo de aprendizaje por refuerzo se trata de una técnica de machine Learning, la cual se ha desarrollado en función al software para tomar las decisiones y mejorar los resultados en el análisis de la información. simula los procesos de aprendizaje mediante ensayos y los errores que los humanos puedan desarrollar al analizar los objetivos. Este tipo de análisis de información no utiliza etiquetas de salida y por lo cual este tipo de análisis no se considera como de tipo supervisado a pesar de que los algoritmos puedan aprender por su cuenta. Cabe destacar que tampoco existe supervisión puesto que no se aplican acercamientos entre las muestras.

El tipo de Machine Learning por refuerzo se diferencia de los dos anteriores debido que se minimiza la función del costo, reduciendo los errores en el análisis de la información, intentando maximizar la recompensa al ofrecer la posibilidad de cometer errores que permiten que el sistema pueda aprender. Los componentes que forman parte de los sistemas de Machine Learning por refuerzo son el agente, el ambiente, la acción, el estado y la recompensa. El agente es el modelo que se desea entrenar con el concepto para tomar las decisiones. El ambiente es el entorno de interacción entre los sistemas. La acción son todas las posibles acciones que se puedan tomar, el estado son los indicadores

del ambiente y la recompensa es la acción tomada por el agente para premiar o penalizar al sistema (Aprende Machine Learning, 2020).

11. Regresión Lineal

Una regresión lineal se caracteriza por ser la técnica de análisis de datos e información, que tiene como objetivo predecir el valor de los datos desconocidos gracias al uso de otro valor de datos según lo relacionado. Todas las regresiones lineales dependen del modelamiento matemático según las variables desconocidas o dependientes y las variables desconocidas e independientes para desarrollar una ecuación lineal. Los modelos de regresión lineal pueden considerarse como simples, pero ofrecen una gran ventaja para mediante el uso de la fórmula matemática lograr la interpretación de los datos y determinar así las predicciones (Amazon, 2022).

Existen tres tipos de regresiones lineales que son la simple, la múltiple y la logística. La regresión lineal simple es un procedimiento que usa una ecuación matemática para expresar la relación entre dos variables. Con ello se logra determinar el valor de una base respecto al valor de otra. Los modelos de regresión simple dependen de dos variables, una dependiente y otra independiente para establecer un funcionamiento sobre sí. Las suposiciones de la regresión lineal simple usan valores de variable X, considerados valores fijos preseleccionados por los investigadores y necesarios en la recolección de datos. Las variables x permiten medir la magnitud de los datos. Además, para cada subpoblación de X en el valor, existirá una distribución normal para los datos Y. todas las varianzas de las subpoblaciones serán consideradas como iguales (Vicerrectoría de Pregrado, 2022). La fórmula de la regresión lineal simple es

$$Y=a+bx$$

Los modelos de regresión lineal múltiple son estadísticos con versatilidad en su composición, permitiendo evaluar las relaciones entre factores continuos y predictores. Todos los predictores pueden considerarse como campos continuos, categóricos y derivados en torno a las relaciones no lineales. Los modelos lineales múltiples poseen factores adictivos en los términos de predictor. Estas regresiones múltiples permiten el desarrollo de gráficos entre dos o más campos para determinar un factor predictor categórico y dos campos continuos auxiliares. En este tipo de modelo se aplican dos o

más variables con influencia de estar correlacionadas con el valor de una tercera variable.

La fórmula es:

$$Y_j = b_0 + b_1x_{1j} + b_2x_{2j} + \dots + b_kx_{kj} + u_j$$

Siendo y la variable endógena, x la variable exógena, el valor u los residuos y los b los coeficientes estimados del efecto marginal (Montero, 2016).

12. KMedias para la separación de grupos de importancia

KMedias o también conocido como KMeans se trata de el desarrollo de un algoritmo que permite una clasificación no supervisada, que tiene la labor de agruparse con distintos objetos en un determinado número de cuadros. La clasificación se desarrolla en base a sus características para que la agrupación sea eficiente, minimizando así también la suma de las distancias existentes entre los distintos objetos. Para el control de los datos, es necesaria la utilización de la función de la distancia cuadrática, la cual se desarrolla mediante la elevación al cuadrado de todas las observaciones para obtener la respectiva media cuadrática y extraer la raíz de la media a fin de dejar una unidad media original (Universo Fórmulas, 2024).

Para la implementación del algoritmo, se puede considerar la utilización de tres determinados pasos como son la iniciación, la asignación de los objetos de los centroides y la actualización de los centroides. Durante el proceso de iniciación, se deberá de escoger un valor en grupos con valor de K , estableciendo así los centroides y el espacio de los datos que han sido elegidos de forma aleatoria. La asignación de los objetos de los centroides deberá estar determinado por el centroide más cercano y con su respectiva actualización, se deberá analizar la posición del centroide tomando en consideración la posición de los objetos pertenecientes en el grupo. Cada uno de estos procesos deberá de repetirse hasta establecer un movimiento de la distancia umbral sobre cada paso (Uni Oviedo, 2024).

La fórmula de KMeans es:

$$J = \sum_{j=1}^k \sum_{i=1}^n \|x_i^{(j)} - c_j\|^2$$

Siendo la J el objetivo de la función, la K el número del clúster, la N el número del caso, la X el caso y la C el centroide del clúster ubicado dentro de su función distancia (Uni Oviedo, 2024). Su funcionamiento permite que la toma de decisión de la aplicación funcione de una forma más eficiente, enfocándose principalmente en la recopilación de la información de grandes cantidades de datos con las cuales se pueden determinar distintos patrones. En el caso de la aplicación para la empresa EDEX Education Group en el manejo de la información, se logra establecer patrones de comportamiento de los clientes, para en base a las predicciones determinadas, se logre comprobar su acercamiento con la empresa a fin de determinar si será un cliente fijo o desertará en un determinado proceso de adquisición del servicio.

El algoritmo garantiza una interacción eficiente para establecer un resultado adecuado en relación al comportamiento del consumidor. El patrón deberá permitir la separación de los grupos de los clientes según su importancia para la empresa, a fin que se puedan priorizar sobre determinados comportamientos en el mercado y definir su impacto para la empresa, permitiendo con ello una clasificación intuitiva sobre los grupos que forman parte del proceso de análisis. Con ello se busca mejorar la elección de los centros iniciales, sobre la realización de los procedimientos y sobre la elección en el proceso de opción en la minimización de los riesgos potenciales (Ramírez, 2023).

Los datos son analizados como el proceso de reordenamiento aleatorio de los datos que han sido escogidos en primera instancia. El reordenamiento aleatorio de los datos es posible gracias al desarrollo del Random Partition, siendo esta acción necesaria para generar una partición aleatoria sobre el conjunto de los datos y sobre su cálculo en sus centroides. Se podría escoger también una propuesta de Furthest Point Heuristic con la que se logre escoger el centroide que va a ir añadido sobre el resto. El centro del clúster se lo escoge sobre los centroides previamente elegidos. El proceso de Sorting Heuristic evidencia el uso de puntos de muestra para ordenar mediante este criterio, mismo que funcionará de forma uniforme y sobre los clústeres bien separados. El sistema Projection Based Heuristics se trata del desarrollo de los puntos sobre los ejes, los cuales podrían ser el ejemplo para la dirección principal y la participación del eje en un determinado número de segmentos, permitiendo el cálculo posterior de cada segmento. Density Based Heuristics se menciona como el cálculo de los puntos según la densidad de cada punto. Y para la iniciación del K Means, es necesario la elección de los centros de manera que los

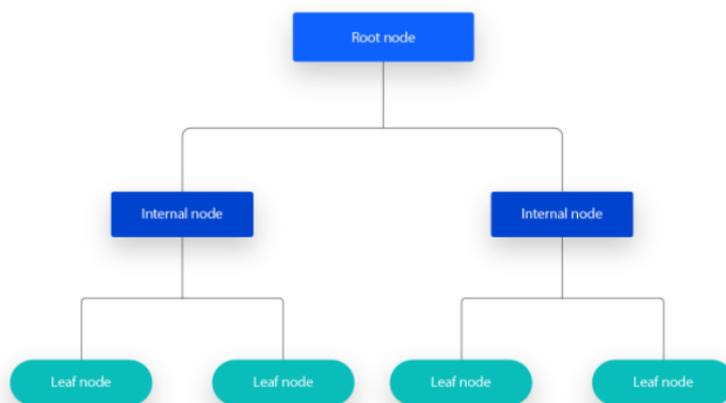
puntos más alejados existentes en los centroides escogidos previamente, puedan disponer de una probabilidad alta para ser seleccionadas (Perucha, 2022).

13. **Árbol de Decisión**

El sistema **Árbol de Decisión** consiste en el desarrollo de mapas aleatorios sobre todos los posibles resultados que se puedan generar en una serie de decisiones relacionadas. Gracias a esta herramienta es posible determinar de forma más efectiva el comportamiento de los usuarios basados un determinado comportamiento. El árbol de decisión tiene relación con el desarrollo del aprendizaje no paramétrico, utilizado en las regresiones y clasificación de los sistemas. El sistema posee además una estructura de árbol de jerarquía, dependiendo de otros elementos como nodos, ramas, nodos intentos y nodos hoja.

El árbol de decisiones posee una característica peculiar y es la de crear nodos de la raíz, misma que no posee ramas entrantes. A medida que se va llenando la información, los nodos van en aumento, evolucionando en los resultados y en la forma que se generan los subconjuntos homogéneos y sobre las hojas terminales, siendo estos resultados posibles dentro de la serie de datos. Es gracias al uso de los diagramas de flujo que se consigue una representación fácil sobre las respectivas tomas de decisiones y sobre los diferentes grupos en las organizaciones que están determinadas por la toma de decisiones.

Figura 5 Esquema del árbol de decisión



Nota. Esquema del árbol de decisión que comienza con un nodo general para abrir más opciones. Recuperado de (IBM, 2024).

Gracias al aprendizaje de árbol, ha sido posible el diseño de estrategias que dividen los elementos existentes para fomentar una mejor toma de decisión, basada en una acción recursiva a fin de determinar determinados patrones. Son los árboles de decisiones los que tienen principio de parsimonia, que se desarrollan más allá de las necesidades en la complejidad de la toma de decisiones. La reducción de la complejidad del sistema se encuentra en la poda, como un proceso que permite ajustar el modelo sobre la evaluación gracias al proceso de validación cruzada (IBM, 2024).

14. Random Forest

El sistema Random Forest, también conocido como bosque aleatorio forma parte del sistema de algoritmos de Machine Learning de uso común, creado por Breiman y Cutler para mejorar el proceso de toma de decisiones en sus resultados. La facilidad de su utilización permite que los usuarios puedan combinar las ideas relacionadas con los árboles en las decisiones para desarrollar un único resultado que será el relacionado con el comportamiento de consumo. Gracias a su facilidad para ser utilizado, es posible que el uso y su flexibilidad permitan mejorar la toma de decisiones para obtener un mejor resultado, generando una importante facilidad en el proceso de toma de decisiones. Gracias al esquema de los árboles de decisión, se consigue una forma eficiente en la que se mejora la toma de decisiones por medio de la utilización de los subconjuntos de los datos, basados en algoritmos para la clasificación de la regresión. El sistema puede utilizar elementos como la impureza de Gini, las ganancias de la información existentes en el error cuadrático y la evaluación en la calidad de la división (Cutler, Richard, & Stevens, 2011).

El método más utilizado para analizar los procesos de aprendizaje es por medio del método de los conjuntos, el cual se compone principalmente de varios conjuntos de elementos que han sido clasificados previamente a fin de identificar de manera adecuada el resultado popular aplicado. El modelo permitió la implementación de un sistema de muestra aleatoria en la que se consigue que se analicen los conjuntos de entrenamiento de reemplazo, sobre los valores de los datos, siendo elegidos nuevamente, permitiendo al sistema entrenar de forma independiente las acciones sobre la regresión, el promedio y las estimaciones de los análisis de datos (IBM, 2024).

El sistema Random Forrest tiene la capacidad de combinar distintos algoritmos considerados como débiles para el aumento del poder predictivo. Los algoritmos que se conocen como débiles pueden desarrollar distintas variables de carácter dependiente, lo que ocasiona que las predicciones sean poco eficientes. El no contar con algoritmos que puedan combinar los distintos modelos de algoritmos genera que el poder de la predicción del algoritmo sea poco eficiente. Entre los algoritmos de poco impacto en el proceso de selección se encuentran el Bagging, el cual es un modelo de algoritmo que funciona de manera independiente y logra combinar los modelos de forma independiente y desarrollar una combinación de resultados posibles. La clasificación se generaría por medio del voto por mayoría como en las estadísticas se obtendrían los valores de la moda. El Random Forest que utiliza un sistema Bagging, está enfocado en mejorar el proceso de toma de decisiones por medio del entrenamiento del sistema de decisión con valores previos, obteniendo así una media para resolver los problemas de regresión existentes (Díaz, 2024).

15. Clustering

El Clustering está relacionado con la formación de grupos para su posterior análisis, siendo esta una tarea que pretende agrupar determinados objetivos sobre una serie de conjuntos de forma que los miembros del mismo equipo podrán determinar su particularidad y parecido con otros grupos de características similares. El Clustering puede ser altamente eficiente para los análisis de patrones, siendo utilizados en el estudio del entorno con el cual se forman grupos homogéneos, o grupos cerrados, con los cuales se puede compartir por medio de una serie de elementos de distintas características, determinada propiedad que mantengan cierta similitud. Todos los datos obtenidos son generados en la Big Data, con el que se adquieren los datos sobre las fuentes para ser preparados sobre las fuentes. El proceso de análisis de la información se genera mediante el uso de los datos preparados para ser analizados sobre el proceso de exploración y trato de la información, con la que se pueden valorar, el valor. El proceso de análisis se selecciona sobre las técnicas adecuadas con las que se puede obtener la información del valor y su regresión, clasificación y asolación del valor dentro del modelo analítico. Se establece un proceso de comunicación sobre los resultados para la construcción de los informes, la visualización de la transmisión de la información y sobre los procesos de análisis aplicados (Lejarza, 2024).

Para considerar el uso de la estrategia Clustering, es necesario que los Clustering tengan calidad en sus datos, se disponga del número total de los grupos que son para el cálculo, que se determinen las diferencias de cada grupo y el tamaño del clúster. Para finalmente, considerar la jerarquía en el momento de definir los grupos. La aplicación de las herramientas Clustering quedan definidas para analizar el comportamiento del mercado por medio de la segmentación de los clientes y sus determinados grupos. Siendo este factor fundamental para ser controlado por parte de la empresa EDEX y determinar el comportamiento de sus clientes. La segmentación del mercado es la forma más efectiva de utilizar el proceso Clustering, siendo este caracterizado para establecer determinadas propiedades que permitan descubrir el modelo del comportamiento y el interés para el ofrecimiento de determinados productos y servicios personalidades sobre los gustos y los intereses (UNIR, 2021).

III. CAPITULO III: MARCO LEGAL

En el ámbito legal se han generado varios planes alineados a una normativa legal entre los cuales se destacan: El Plan Nacional de Telecomunicaciones y Tecnologías de Información del Ecuador 2016-2021. El Convenio Marco de Cooperación Interinstitucional firmado con los diferentes entes como el Ministerio de Telecomunicaciones y de La Sociedad de la Información, la Secretaría de Educación Superior, Ciencia, Tecnología e Innovación. Los Planes Nacionales para el Desarrollo (2013-2017). La Ley Orgánica de Telecomunicaciones. El Código Orgánico de La Economía Social de los Conocimientos, Creatividad e Innovación.

Así como los artículos que contemplan el desarrollo de la investigación científica y de la innovación tecnológica, de las TIC y diversas áreas del conocimiento que están en la Constitución de la República del Ecuador de los cuales se puede destacar los (Art. 281, Art. 385, Art. 387, Art. 423, Art. 388, Art. 313) que permiten alinearse y ser partícipes de la adopción de estas tecnologías emergentes para el cambio y adopción de una cultura de transformación digital para un mejor desarrollo a nivel empresarial en el ámbito local con proyección internacional, lo que se convierte en un impacto positivo

IV. CAPÍTULO IV: METODOLOGIA

1. Etapa 1: Implementación de los algoritmos K Medias para la toma de decisiones

Edex Education Group nos proporcionó la base de datos limpia y concisa, por lo que no se tuvo que hacer un proceso de data cleaning.

Para esta investigación se va a emplear los algoritmos de K medias, Árbol de decisión y Random Forest para la clasificación. K means es un algoritmo que se usa para clasificar y agrupar objetos basados en sus propias características. Este método no supervisado lleva a cabo la segmentación en base a la similitud de sus características. Nos permite analizar una gran cantidad de datos y arroja buenos resultados.

El segundo algoritmo que usaremos en este proyecto es árbol de decisión es un algoritmo que se utiliza para la regresión y para la clasificación de datos, este algoritmo tiene como objetivo evaluar las variables para así poder seleccionar el mejor, su nombre es debido a su estructura que es semejante a un árbol pero invertido, permite evaluar mediante un gráfico los posibles resultados y consecuencias de una decisión en particular.

Por último, usaremos el algoritmo Random Forest que nos sirve para clasificar dentro de una categoría específica combinando múltiples modelos para mejorar el rendimiento.

2. Etapa 2: Pasos a desarrollar

Para la creación del modelo se va a usar RStudio, el cual es un entorno de desarrollo de software estadístico en el que se permite trabajar y manejar con el lenguaje de programación R. Para poder realizar el modelo primero hay que llamar a la base de datos con la función (read.csv) o (read.csv2) dependiendo del tipo de base a utilizar.

Una vez que tengamos la base de datos preparada, procedemos a programar las funciones necesarias para implementar Kmedias para separar por grupos de importancia, para esto debemos descargar una serie de paquetes que detallaremos a continuación:

Una vez que la base de datos esta lista y los paquetes y librerías necesarios instalados correctamente, el siguiente paso será identificar a los alumnos que pueden viajar y a los que no, (Standard y Premium), aplicando el método de K-medias para separar por grupos de importancia. Antes de aplicar el algoritmo K-medias hay que escalar los datos para poder observar todas las variables que tengan la misma escala, se realiza con la función (SCALE), aplicada al conjunto de datos “study”.

```
install.packages("tidyverse")
```

```
library(tidyverse)
```

```
library(cluster)
```

```
library(factoextra),
```

```
install.packages("NbClust")
```

```
library(NbClust)
```

El “tidyverse” es una colección de paquetes que se pueden instalar fácilmente con un solo "meta"-paquete. Esto ofrece una forma fácil y rápida de descargar e instalar todos los paquetes “tidyverse” utilizando el comando R. Los paquetes que probablemente utilice en el análisis de datos cotidianos se adjuntan al núcleo “tidyverse”.

La primera librería que se va a usar es “Clúster”, que es la abreviación de “Finding Groups in Data” el cual proporciona las herramientas necesarias para analizar clúster y agrupación de datos. Es una herramienta muy útil a la hora del análisis de clustering en RStudio, nos ayuda a identificar si es que hay grupos similares de datos y evaluar su calidad, ya que eso nos va ayudar a obtener mejores resultados.

Procedemos a implementar la librería “factorextra” que es una abreviación de “Extract and Visualize the results of multivariate data analyses”, que nos proporciona algunas funciones fáciles de usar para extraer y visualizar los resultados que arroja el análisis de datos multivariantes de los distintos paquetes de R. Además, contiene funciones para simplificar algunos pasos del análisis de agrupaciones y proporciona una mejor visualización de datos en ggplot2.

#Escalamos

```
SKStudy<- scale(Study)
```

Es una función genérica cuyo método predeterminado centra o escala las columnas de una matriz numérica.

El proceso de escalado de las columnas (después del centrado) está determinado por el valor de escala. Si el vector de tipo numérico escala tiene una longitud igual al número de columnas x , cada columna de x se divide por el valor de escala correspondiente. Si el escalado es TRUE, el escalado se realiza dividiendo las columnas (centradas) de x por sus desviaciones estándar; en caso contrario, se divide por la raíz cuadrática media. El escalado no se realiza si la escala es FALSE

A partir de la estimación del número de clústeres con el paquete “NbClust” se podrá estimar el número óptimo de clústeres. Un rango considerable de distancia es de 2 a 10 clústeres, así forzamos al algoritmo a no salirse de ese rango porque luego se empiezan a formar otro conjunto de datos. La distancia euclidiana es una medida de distancia entre dos puntos en específico para medir la similitud entre puntos de datos. El método implementado es K-Medias para estimar el número de clúster.

#Estimar el número de cluster

```
NumClust<-NbClust(SKStudy, distance = "euclidean", min.nc=2, max.nc=10,  
method = "kmeans", index = "alllong")
```

“NbClust” que es la abreviación de “Determining the best number of clusters in a data set” tiene el objetivo de determinar el número de conglomerados e indicarnos cuál es el mejor esquema de agrupamiento en base a los resultados obtenidos. Proporciona una serie de métodos para evaluar la calidad de las particiones de las técnicas de clustering. Nos permite realizar análisis comparativos entre diferentes clústeres y así poder determinar de manera sencilla y organizada cuál es el número óptimo de clústeres necesarios para este desarrollo en particular.

A continuación, se calcularán los clústeres, para esto utilizaremos el paquete (Kmeans) para realizar el algoritmo de K-medias en los datos previamente escalados. Se desean encontrar 2 centros de clústeres.

```
#calculamos los dos clústeres
```

```
k2 <- kmeans(SKStudy, centers = 2, nstart = 25) k2 str(k2)
```

A partir de esto se van a plotear los clústeres, para poder visualizar los resultados utilizando el modelo de K-medias (k2) en los datos escalados. De igual manera que ploteados los clústeres, en esta ocasión con elipse configurada en “euclid” para visualizar elipses confiables se usa “repel”=TRUE” (para evitar superposiciones de etiquetas). La herramienta “star.plot=TRUE” agrega un gráfico de estrella para evidenciar la varianza realizada. Se plotea con elipse normal para visualizar las elipses que se encuentran alrededor de los clústeres.

```
#plotear los cluster
```

```
fviz_cluster(k2, data = SKStudy),fviz_cluster(k2, data = SKStudy, ellipse.type =  
"euclid",repel = TRUE,star.plot = TRUE) #ellipse.type= "t", "norm", "euclid"  
fviz_cluster(k2, data = SKStudy, ellipse.type = "norm")
```

Con este resultado procedemos a pasar los clúster a la base de datos.

```
#Pasar Cluster a la base
```

```
Study$K_Means<-as.factor(k2$cluster) table(Study$K_Means)
```

Como siguiente paso se realizará la codificación de la variable de respuesta con la función (factor) que nos ayuda a proporcionar varias visualizaciones de los clústeres gracias al análisis de K-Medias, esto nos ofrece varios puntos de vista sobre la distribución y los grupos identificados mediante los datos escalados.

```
#Codifico la variable de respuesta
```

```
Study$K_Means <- factor(Study$K_Means, levels = c(1,2), labels =  
c("Estandar","Primium"))
```

```
Study$Funcion_Propietario <- factor(Study$Funcion_Propietario, levels =  
c(1,2,3,4,5),  
labels = c("Consultor Ecu ES", "Consultor Ecu Idiomas", "Coordinador",  
"Director", "Operador"))  
  
Study$Nivel_educativo <- factor(Study$Nivel_educativo, levels = c(1,2), labels  
= c("Bachillerato", "Tercer nivel"))
```

Con la semilla establecida se marca el inicio para crear nuestro árbol de decisión.
Procedemos a preparar los datos de entrenamiento.

```
#Creamos arbol  
  
library(caret)  
  
#para un arbol o cualquier derivado de un arbol de decision, se planta una  
semilla  
  
set.seed(123)
```

Consecutivamente instalamos la librería “Caret” que es la abreviatura de “Classification and Regression Training”, nos permite agilizar el entrenamiento del modelo de clasificación y regresión de datos. Nos ofrece una conexión que simplifica el proceso de un modelo empleado a la mayoría de los métodos utilizados en RStudio para realizar tareas comunes en el desarrollo de los modelos de aprendizaje automático independientemente del algoritmo que se esté utilizando.

```
#entrenamos los datos  
  
dataentrenamiento <- createDataPartition(Study$K_Means, p=0.80, list =  
FALSE) library(randomForest)
```

Cuando hablamos de crear particiones de datos, nos referimos a dividir en un conjunto de datos en subconjuntos para un objetivo en específico. En este caso “createdatapartition” toma como argumento la variable “Study” que es lo que queremos

clasificar y la proporción deseada para el conjunto de entrenamiento. Luego “partition” es un vector que nos ayuda a dividir en conjuntos de datos de entrenamiento y prueba.

Con esto procedemos a construir el bosque aleatorio con la función de la librería (randomforest), en el mismo tenemos que especificar las filas correspondientes al conjunto de entrenamiento (x=study(dataentrenamiento, 1:6), luego se selecciona la columna 7 de la base de datos como la variable de respuesta, seguido se especifica el número de árboles en el bosque aleatorio (3000) para así lograr la predicción.

#Paso 4.- Armamos el bosque aleatorio

```
RandomTreeModel <- randomForest(x=Study[dataentrenamiento, 1:6],  
y=Study[dataentrenamiento,7], ntree = 3000, keep.forest = TRUE)  
RandomTreeModel$importance
```

#Paso 5.- Realizo la prediccion

```
Prediccion <- predict(RandomTreeModel,Study)
```

Prediccion

Como paso final, se evalúa el rendimiento del modelo realizado del bosque aleatorio implementando la matriz de confusión para así realizar predicciones totales en el conjunto de datos que es utilizado como un algoritmo de aprendizaje supervisado que se utiliza tanto para tareas de clasificación de regresión. Es conocido por la capacidad de manejar grandes conjuntos de datos con muchas variables predictorias, de clasificación y regresión basadas en un bosque de árboles que utilizan entradas aleatorias.

#Paso 6.- Armo la matriz de confusion

```
Matriz <- table(Study$K_Means, Prediccion, dnn = c("Actual", "Predicho"))  
Matriz
```

#Valores clasificados

```
Study$Prediccion <- predict(RandomTreeModel,Study)
```

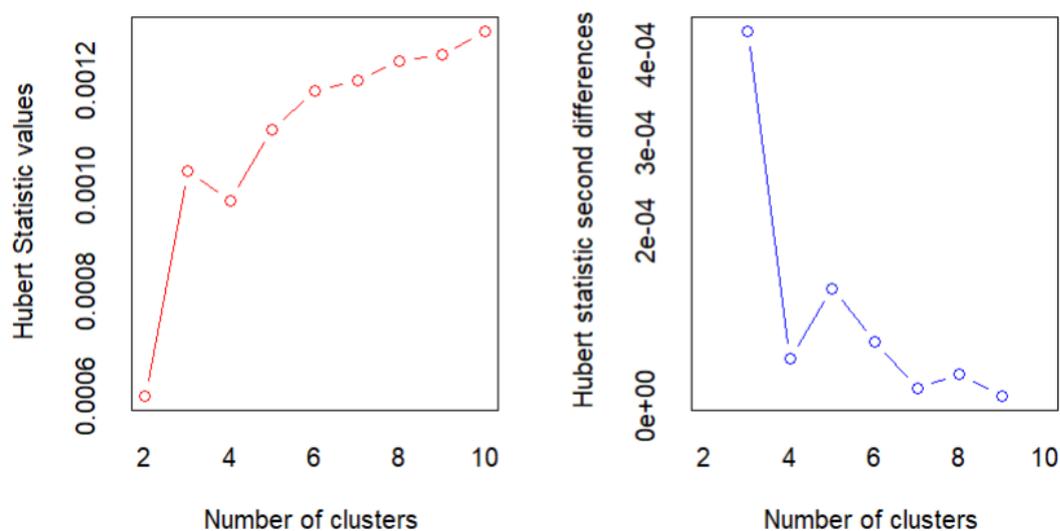
#Probabilidad de exito

```
Study$RandProb <- predict(RandomTreeModel,Study, type = "prob")
```

V. CAPÍTULO V: RESULTADOS

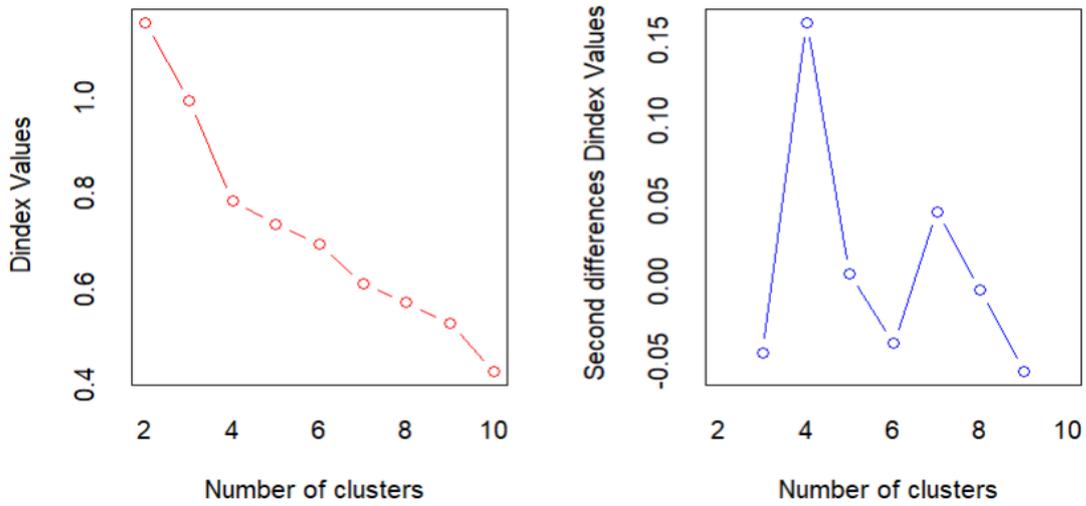
A través de la aplicación de los modelos previamente empleados en el marco de esta investigación, hemos conseguido llevar a cabo un estudio orientado a determinar la audiencia a la cual la empresa debería dirigirse al ofrecer cursos en el extranjero. Inicialmente, empleamos el paquete *clúster*, permitiéndonos la agrupación de variables con el fin de que los datos sean homogéneos.

Figura 6 Gráfico de clústeres



Nota. Obtenido de RStudio a partir de la codificación escrita por autor

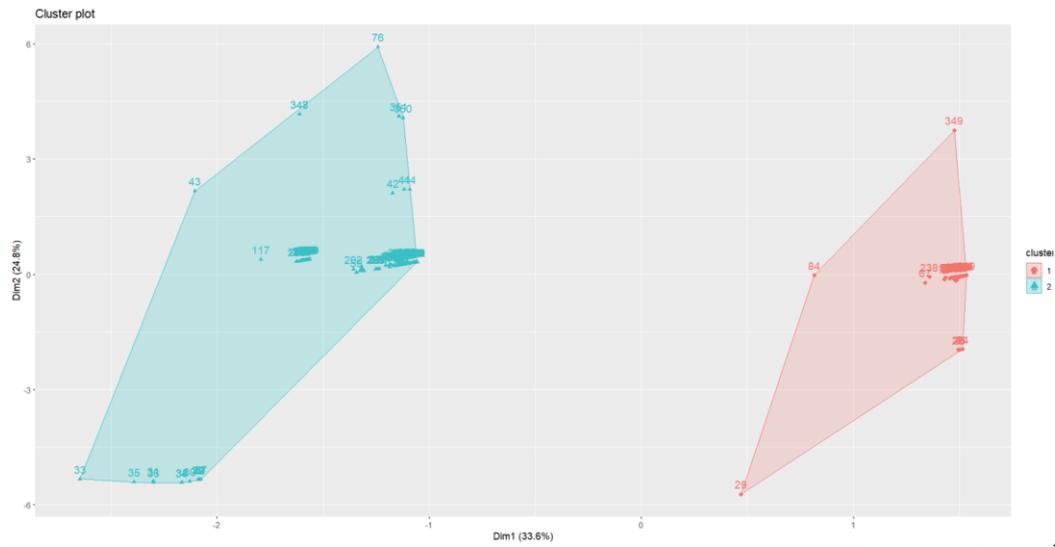
Figura 7 Gráfico de clústeres



Nota. Obtenido de RStudio a partir de la codificación escrita por autor

En las figuras 6 y 7, se representan los gráficos de clústeres, donde se identifica que, en ambas simulaciones, el punto óptimo de los clústeres es entre 3 y 4 clústeres para la estructuración de los datos.

Figura 8 Gráfico de clústeres conglomerados



Nota.

Obtenido de RStudio a partir de la codificación escrita por autor

Se puede observar en el gráfico 8 que en el grupo 2 (premium) existe una mayor heterogeneidad entre los grupos, es decir que a este conjunto se lo puede llamar a muchos campos de estudio y no tienen la incógnita de no saber que hacer, mientras que el grupo 1 (estándar) no tiene mucha disgregación, lo que significa que necesitan más ayuda para decidir qué tipo de estudio cursar.

Figura 9 Gráfico elíptico tipo euclidiana



Nota. Obtenido de RStudio a partir de la codificación escrita por autor

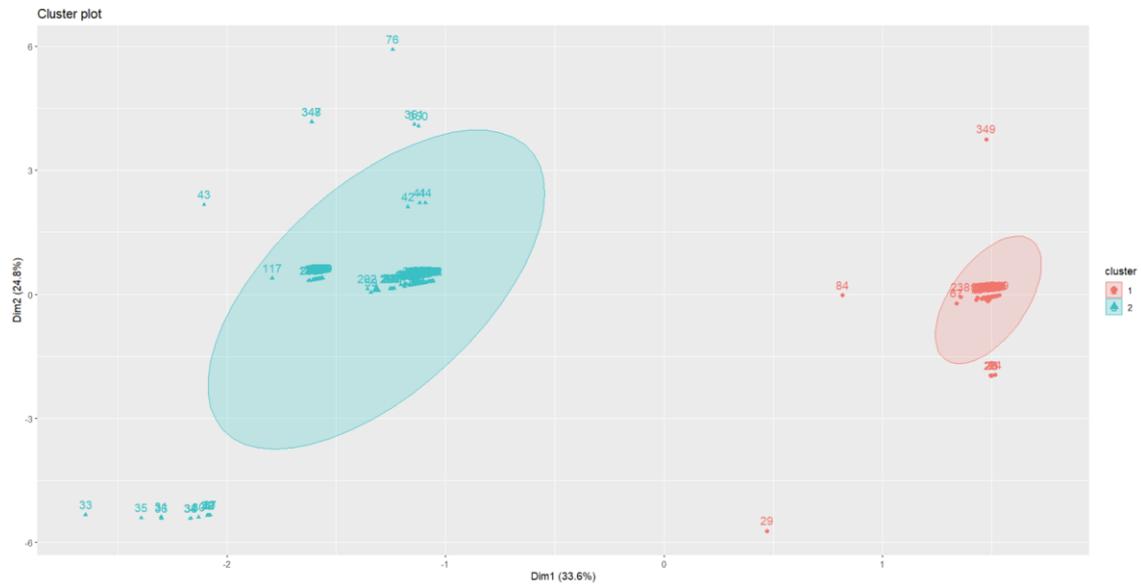
El análisis de la Figura 9 revela que en ambos conjuntos de datos existe un grupo significativo de personas que no se ajustan a la agrupación predominante. Esto sugiere la posibilidad de que estas personas puedan formar parte de un conjunto de datos distinto, cuyo direccionamiento o comportamiento no ha sido identificado debido a que aún no está claramente definido.

Al observar las tendencias, se evidencia la posibilidad de esta probabilidad. Esto se respalda al examinar las dimensiones de cada clúster, lo que nos proporciona una varianza explicada del 58,4%. Esta cifra representa la cantidad de variabilidad en los datos que puede ser explicada por la agrupación realizada. Una varianza explicada del 58,4% sugiere que existe una probabilidad significativa de que los datos sigan saliendo del conjunto de manera no homogénea.

El propósito detrás de este análisis es demostrar que es necesario hacer que los conjuntos de datos sean más homogéneos. Esto se puede lograr mediante la comunicación directa con las personas involucradas, lo que permitiría comprender mejor sus características, preferencias y comportamientos. Esta comprensión más profunda

ayudaría a dirigir a las personas hacia campos específicos que puedan ser más relevantes para ellos, mejorando así la eficacia de cualquier estrategia o dirección que la empresa o entidad esté considerando.

Figura 10 Gráfico de aplicación de k-means en los clústeres



Nota. Obtenido de RStudio a partir de la codificación escrita por autor

Una vez optimizados los clúster, podemos ver cuántas personas se quedaron fuera de los conjuntos, en el clúster 2 hay más diversificación de datos ya que dichas personas tienen más opciones a escoger.

Tabla 1 Nivel de importancia

	MeanDecreaseGini
Funcion_Propietario	0.4224111
Inversion	7.9514898
Antiguedad	0.4913578
Nivel_educativo	0.4734066
Solvencia	65.4976472
obligatoria	64.1864229

Nota. Obtenido de RStudio a partir de la codificación escrita por autor

En la Tabla 1, se muestra la relevancia de las variables en el modelo de clasificación RandomForest. Esto revela la contribución de cada variable en la reducción de la impureza en los nodos del árbol de decisión. Sin lugar a dudas, la solvencia económica destaca como un factor crucial para el modelo, evidenciando una importancia cercana al 65.497, ya que sin la solvencia el estudiante no podría presentarse ante la

embajada del país de destino de su proyecto de estudio, los otros factores no son un impedimento para que el estudiante empiece sus estudios en el extranjero.

Tabla 2 Matriz de Confusión

	Predicho	
Actual	Estandar	Primum
Estandar	186	0
Primum	0	165

Nota. Obtenido de RStudio a partir de la codificación escrita por autor

La finalidad de la matriz de confusión es proporcionar una visión detallada del rendimiento del modelo de clasificación RandomForest. A partir de los valores en la matriz de confusión podemos establecer que no existen errores en la clasificación inicial.

Tabla 3 Tabla de frecuencia por nivel educativo

	Var1	Freq	Fre1	FreAcu	FrReAc
1	Estándar	186	0.53	186	0.53
2	Primum	165	0.47	351	1.00

Nota. Obtenido de RStudio a partir de la codificación escrita por autor

En la tabla número 3, podemos evidenciar la probabilidad de obtener un estudiante estándar con un 53% versus un estudiante Premium con un 47%

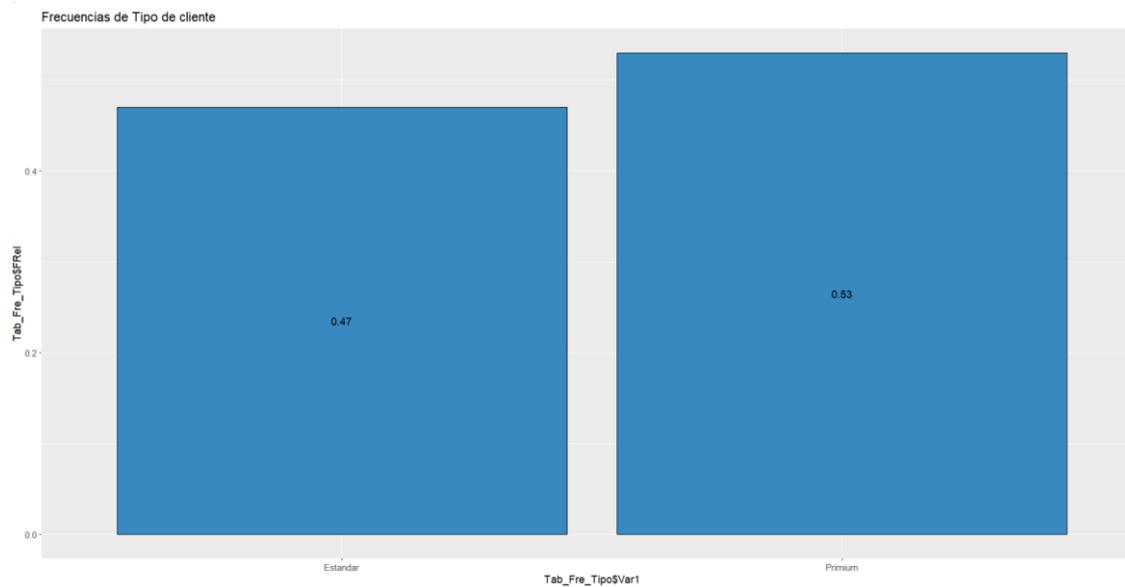
Figura 11 Predicciones a través de K-means y RandomForest

	Funcion_Propietario	Inversion	Antiguedad	Nivel_educativo	Solvencia	Obligatoria	K_Means	Prediccion	RandProb[, "Estandar"]	RandProb[, "Primum"]
1	Consultor Ecu Idiomas	954.6192	0	Bachillerato	16794	2	Primum	Primum	0.0003333333	0.9996666667
2	Consultor Ecu Idiomas	2231.0700	0	Bachillerato	12000	1	Estandar	Estandar	0.9986666667	0.0013333333
3	Consultor Ecu Idiomas	1784.8560	0	Bachillerato	12000	1	Estandar	Estandar	0.9986666667	0.0013333333
4	Consultor Ecu Idiomas	930.7537	0	Bachillerato	16794	2	Primum	Primum	0.0003333333	0.9996666667
5	Consultor Ecu Idiomas	1829.6868	0	Bachillerato	16794	2	Primum	Primum	0.0086666667	0.9913333333
6	Consultor Ecu Idiomas	2278.4573	0	Bachillerato	16794	2	Primum	Primum	0.0106666667	0.9893333333
7	Consultor Ecu Idiomas	662.9300	0	Bachillerato	11800	1	Estandar	Estandar	0.9990000000	0.0010000000
8	Consultor Ecu Idiomas	499.7597	0	Bachillerato	12000	1	Estandar	Estandar	0.9983333333	0.0016666667
9	Consultor Ecu Idiomas	1204.7778	1	Bachillerato	12000	1	Estandar	Estandar	0.9966666667	0.0033333333
10	Consultor Ecu Idiomas	954.6192	0	Bachillerato	16794	2	Primum	Primum	0.0003333333	0.9996666667
11	Consultor Ecu Idiomas	2007.9630	0	Bachillerato	12000	1	Estandar	Estandar	0.9986666667	0.0013333333
12	Consultor Ecu Idiomas	348.0469	0	Bachillerato	12000	1	Estandar	Estandar	0.9990000000	0.0010000000
13	Consultor Ecu Idiomas	212.6953	0	Bachillerato	12000	1	Estandar	Estandar	0.9990000000	0.0010000000
14	Consultor Ecu Idiomas	464.0626	0	Bachillerato	12000	1	Estandar	Estandar	0.9990000000	0.0010000000
15	Consultor Ecu Idiomas	636.4128	0	Bachillerato	16794	2	Primum	Primum	0.0150000000	0.9850000000
16	Consultor Ecu Idiomas	1312.6014	0	Bachillerato	16794	2	Primum	Primum	0.0003333333	0.9996666667
17	Consultor Ecu Idiomas	1193.2740	1	Bachillerato	16794	2	Primum	Primum	0.0016666667	0.9983333333
18	Consultor Ecu Idiomas	2007.9630	7	Bachillerato	12000	1	Estandar	Estandar	0.9986666667	0.0013333333
19	Consultor Ecu Idiomas	1695.6132	7	Bachillerato	12000	1	Estandar	Estandar	0.9986666667	0.0013333333
20	Consultor Ecu Idiomas	763.2000	0	Bachillerato	10000	1	Estandar	Estandar	0.9980000000	0.0020000000
21	Consultor Ecu Idiomas	428.4000	0	Bachillerato	10000	1	Estandar	Estandar	0.9990000000	0.0010000000
22	Consultor Ecu Idiomas	428.4000	0	Bachillerato	10000	1	Estandar	Estandar	0.9990000000	0.0010000000
23	Consultor Ecu Idiomas	345.0000	0	Bachillerato	10000	1	Estandar	Estandar	0.9990000000	0.0010000000
24	Consultor Ecu ES	795.5160	0	Bachillerato	16794	2	Primum	Primum	0.0040000000	0.9960000000
25	Consultor Ecu ES	1050.0811	0	Bachillerato	16794	2	Primum	Primum	0.0010000000	0.9990000000

Nota. Obtenido de RStudio a partir de la codificación escrita por autor

Como se observa en la tabla número 3, se encuentra un modelo de negocio mejor al que se está empleando en la actualidad

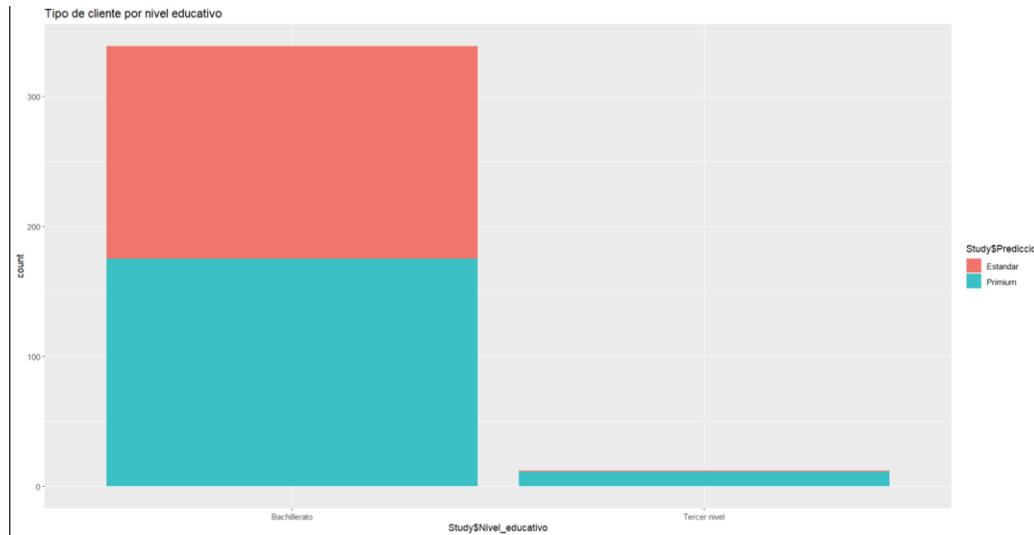
Figura 12 Frecuencia tipo de cliente



Nota. Obtenido de RStudio a partir de la codificación escrita por autor

Observando la figura número 12, podemos establecer que el conjunto premium ya tiene establecido en que área se quiere especializar, en cambio el grupo Estándar tiene que vincularse al curso que quiere cursar y tiene que ser instruido por su asesor.

Figura 13 Tipo de cliente por nivel educativo



Nota. Obtenido de RStudio a partir de la codificación escrita por autor

El análisis final indica que el conjunto de datos correspondiente al Bachillerato, la mayor parte de los estudiantes muestran una tendencia hacia la indecisión en comparación al grupo Premium, donde ya tienen establecido que certificación o curso elegir. Este grafico es básicamente una ayuda para el conteo de riesgo, lo que implica que ayuda a la identificación de las áreas donde existe una mayor incertidumbre. Este procedimiento posibilitó la identificación de patrones y la categorización eficiente de segmentos específicos dentro de la población estudiada, proporcionando así una base analítica robusta para la posterior toma de decisiones estratégicas en la oferta de cursos internacionales. Este planteamiento ayuda a que la empresa adapte sus programas de manera mas precisa para asi poder abordar las necesidades y preferencias de cada grupo de estudiantes.

VI. CAPÍTULO VI: CONCLUSIÓN

1. Conclusión

Hoy en día los modelos de Machine Learning son muy importantes en el ámbito empresarial si es que la empresa quiere destacar entre sus competidores y clientes, implementar este concepto brinda muchos beneficios que pueden ser muy atractivos para una empresa y estos son los modelos de autoaprendizaje que amplían la visión y aumentan las probabilidades de tomar decisiones más exactas.

En base a todo el análisis es crucial que los asesores dirijan su atención hacia el grupo Estándar que son los grupos que están cursando sus últimos años de colegio ya que se ha observado un cambio significativo en la dinámica del negocio educativo. Ya que necesitan una guía según sus aptitudes y gustos, a diferencia del grupo Premium. Actualmente, el enfoque principal se ha desplazado hacia este grupo en particular, lo que implica una mayor demanda y atención por parte de las instituciones.

Este cambio se debe a la creciente necesidad de proporcionar una educación de calidad y accesible a un público más amplio. Mientras que el grupo Premium sigue siendo importante, la tendencia indica que, a medida que fortalecemos nuestra posición en el grupo Estándar, la atracción y la transición hacia el grupo premium ocurrirán de manera más natural y automática, brindando oportunidades de crecimiento sostenible a largo plazo.

Se logró establecer que el Machine Learning es una disciplina tecnológica proveniente del uso de la Inteligencia Artificial, que está enfocada en el análisis de los datos, mejorar el proceso de toma de decisiones. El aprendizaje computacional garantiza que las computadoras se conviertan en una herramienta de datos que logre mejorar la experiencia de análisis de la información, permitiendo que, por medio de las estadísticas, empresas como EDEX Education Group puedan comprender las variantes existentes en el mercado.

El uso de la metodología Machine Learning está orientada a mejorar los procesos de análisis de datos, automatizando la construcción de los modelos analíticos para que la empresa EDEX Education Group pueda desarrollar estrategias de aprendizaje de datos

basados en la idea de los sistemas para determinar patrones y mejorar la toma de decisiones. En el caso puntual del negocio, el conocer las necesidades de mercado y desarrollar propuestas para captar la atención de los posibles consumidores que tengan tendencia a retirarse del servicio, ofreciéndoles beneficios adicionales que se adopten a sus necesidades de consumo y evitar la pérdida de los clientes.

De igual Manera se quiere acotar un caso similiar en el que se implementó Machine Learning.

La empresa mundialmente conocida, Netflix utiliza la inteligencia artificial para personalizar sus recomendaciones. Su algoritmo de recomendación se basa en Machine Learning que analiza las distintas variables como, búsqueda, clicks, calificaciones de usuario para así poder determinar sus preferencias y poder ofrecerles contenido que les pueda interesar a cada usuario.

El sistema utiliza varias técnicas de la inteligencia artificial como la minería de datos, detección de patrones y clasificación automática para procesar grandes cantidades de información y crearle un perfil a cada usuario. Además utiliza técnicas de procesamiento de lenguaje natural para analizar las descripciones de las películas y series para determinar su tono, género, temática, etc. Esto permite que la plataforma pueda recomendar contenido basado en los gustos de cada usuario.

Con este implemento, la plataforma ha logrado fidelizar a varias persona, reducir su tiempo de búsqueda y aumentar el tiempo que pasan dentro de la platadorma y poco a poco irse perfeccionando creando nuevos exitos y entretener a su audiencia.

BIBLIOGRAFÍA

- Actuaría Asesoramiento Estratégico. (12 de Octubre de 2020). *Modelos predictivos para evitar la deserción de clientes*. Obtenido de <https://actuaría.com.ec/es/modelos-predictivos-para-evitar-la-desercion-de-clientes/>
- Altertecnica. (1 de octubre de 2023). *¿QUÉ ES EL MACHINE LEARNING Y CÓMO USARLO EN TU EMPRESA?* Obtenido de <https://altertecnica.com/que-es-machine-learning-usarlo-empresa/>
- Amazon. (1 de Enero de 2022). *¿Qué es la regresión lineal?* Obtenido de <https://aws.amazon.com/es/what-is/linear-regression/>
- Amazon. (24 de Octubre de 2023). *¿Qué es el aprendizaje automático?* Obtenido de <https://aws.amazon.com/es/what-is/machine-learning/>
- Amazon. (13 de Noviembre de 2023). *¿Qué es el aprendizaje automático?* Obtenido de <https://aws.amazon.com/es/what-is/machine-learning/>
- Aprende IA. (14 de Noviembre de 2023). *Aprendizaje Supervisado*. Obtenido de <https://aprendeia.com/todo-sobre-aprendizaje-supervisado-en-machine-learning/>
- Aprende Machine Learning. (24 de Diciembre de 2020). *Aprendizaje por Refuerzo*. Obtenido de <https://www.aprendemachinelearning.com/aprendizaje-por-refuerzo/>
- Arranz, E. (5 de Mayo de 2022). *¿Qué es el mercado laboral?* Obtenido de <https://fundacionadecco.org/blog/que-es-el-mercado-laboral/#:~:text=El%20mercado%20de%20trabajo%20es,de%20paro%20de%20dicha%20regi%C3%B3n.>
- ASOMIF. (17 de Julio de 2022). *Inteligencia artificial: sólo el 12% de las empresas hoy la usa para superar a la competencia*. Obtenido de <https://asomifecuador.com/2022/07/21/inteligencia-artificial-solo-el-12-de-las-empresas-hoy-la-usa-para-superar-a-la-competencia/>
- Brita Inteligencia Artificial. (25 de Octubre de 2023). *Cómo utilizar el Machine Learning para la segmentación de clientes*. Obtenido de <https://brita.mx/como-utilizar-el-machine-learning-para-la-segmentacion-de-clientes>
- Carlos de la Hoz Academia. (29 de Septiembre de 2019). *La importancia de la formación académica personal para el desarrollo de la sociedad*. Obtenido de

<https://academiadelahoz.com/la-importancia-de-la-formacion-academica-personal-para-el-desarrollo-de-la-sociedad/>

Colomé, R., & Navajas, E. (1 de Enero de 2021). *Efectos de la inteligencia artificial en la economía y el análisis económico*. Obtenido de https://anceargentina.org/download/publicaciones/INTELIGENCIA_ARTIFICIAL_ANCE2021.pdf

Cutler, A., Richard, D., & Stevens, J. (1 de Enero de 2011). *Random Forests*. Obtenido de https://www.researchgate.net/publication/236952762_Random_Forests

Diario El País. (3 de Julio de 2019). *El 30% de los universitarios no encuentra trabajo cuatro años después de graduars*. Obtenido de https://elpais.com/sociedad/2019/07/03/actualidad/1562170185_829639.html

Díaz, R. (16 de Enero de 2024). *Random Forest – Bagging y árboles de decisión*. Obtenido de <https://www.themachinelearners.com/random-forest-python/>

DispatchTrack. (22 de Enero de 2019). *Tipos de Inteligencia Artificial: características y alcances*. Obtenido de <https://www.beetrack.com/es/blog/tipos-de-inteligencia-artificial>

EDEX Education Group . (14 de Noviembre de 2023). *Encuentra tu programa ideal de estudios en el extranjero*. Obtenido de <https://edex.world/>

EDEX Education Group. (23 de Octubre de 2023). *EDEX Education Group*. Obtenido de <https://edexeducationgroup.com/nosotros/>

EDEX Education Group. (23 de Noviembre de 2023). *EDEX Términos y Condiciones*. Obtenido de <https://edexeducationgroup.com/terminosycondiciones/>

EDEX Education Group. (23 de Octubre de 2023). *TÉRMINOS Y CONDICIONES PROMO AUSTRALIA SEPTIEMBRE*. Obtenido de <https://edexeducationgroup.com/terminos-y-condiciones-promo-australia-septiembre/>

Editorial RSYs. (16 de Febrero de 2022). *Importancia de contar con experiencia laboral durante la carrera*. Obtenido de <https://responsabilidadsocial.net/importancia-de-contar-con-experiencia-laboral-durante-la-carrera/>

EF Ecuador. (23 de Octubre de 2023). *Viajes de inglés*. Obtenido de <https://www.ef.com.ec/pg/viajes-idiomas/ingles/>

- EMIS. (10 de Abril de 2023). *EDEX EDUCATION - GROUP ECUADOR S.A.S. (ECUADOR)*. Obtenido de https://www.emis.com/php/company-profile/EC/Edex_Education_-_Group_Ecuador_SAS_es_13762547.html
- Ferrovial. (13 de Noviembre de 2023). *¿Qué es la inteligencia artificial (IA)?* Obtenido de <https://www.ferrovial.com/es/recursos/inteligencia-artificial/>
- Fondo Monetario Internacional. (1 de Abril de 2020). *La globalización: ¿Amenaza u oportunidad?* Obtenido de <https://www.imf.org/external/np/exr/ib/2000/esl/041200s.htm>
- Fundación Aquae. (12 de Noviembre de 2023). *El impacto económico de la Inteligencia Artificial*. Obtenido de <https://www.fundacionaquae.org/wiki/impacto-economico-la-inteligencia-artificial/>
- Golpe, V. (1 de Enero de 2015). *¿Determina para las organizaciones la contratación de jóvenes recién graduados y les acarrea repercusiones?* Obtenido de <https://www.colibri.udelar.edu.uy/jspui/bitstream/20.500.12008/7785/1/Luz%2C%20Valentina%20da.pdf>
- Grupo Atico34. (12 de Noviembre de 2023). *Inteligencia artificial: Definición, tipos y aplicaciones*. Obtenido de <https://protecciondatos-lopd.com/empresas/inteligencia-artificial/#:~:text=Las%20m%C3%A1quinas%20reactivas%20son%20el,y%20son%20incapaces%20de%20evolucionar.>
- Hinestroza, D. (1 de Enero de 2018). *EL MACHINE LEARNING A TRAVÉS DE LOS TIEMPOS, Y LOS APORTES A LA HUMANIDAD*. Obtenido de <https://repository.unilibre.edu.co/bitstream/handle/10901/17289/EL%20MACHINE%20LEARNING.pdf?sequence=1>
- IBM. (17 de Enero de 2024). *¿Qué es un árbol de decisión?* Obtenido de <https://www.ibm.com/es-es/topics/decision-trees>
- IBM. (17 de Enero de 2024). *¿Qué es un bosque aleatorio?* Obtenido de <https://www.ibm.com/mx-es/topics/random-forest>
- Larsen, K., & Becker, D. (1 de November de 2021). *Automated Machine Learning For Business*. New York: Oxford University Press. Obtenido de <https://books.google.com.ec/books?hl=en&lr=&id=d-krEAAQBAJ&oi=fnd&pg=PP1&dq=machine+learning+in+business&ots=iisdYzWaiu&si>

g=SRn4sDxVHT_NbusVwdzXTn9S_ho&redir_esc=y#v=onepage&q=machine%20learnin
g%20in%20business&f=false

Lejarza. (17 de Enero de 2024). *Random Forests*. Obtenido de
<https://www.uv.es/mlejarza/datamine/T4.pdf>

López, D. (3 de Mayo de 2023). *Deserción de clientes*. Obtenido de
<https://economipedia.com/definiciones/desercion-de-clientes.html>

Luna, J. (8 de Febrero de 2018). *Tipos de aprendizaje automático*. Obtenido de
<https://medium.com/soldai/tipos-de-aprendizaje-autom%C3%A1tico-6413e3c615e2>

Maisueche, A. (1 de Septiembre de 2019). *Utilización del Machine Learning en la Industria 4.0*.
Obtenido de <https://core.ac.uk/download/pdf/228074134.pdf>

Medina, E. (25 de Mayo de 2023). *Las 10 empresas líderes en inteligencia artificial y machine learning en 2023*. Obtenido de <https://cronica.tech/tecnologia/software/las-empresas-lideres-en-inteligencia-artificial-y-machine-learning-en-2023/#:~:text=IBM%2C%20SAS%2C%20Datarobot%2C%20Microsoft,las%20soluciones%20de%20cada%20empresa.>

Montero, R. (1 de Diciembre de 2016). *Modelos de regresión lineal múltiple*. Obtenido de
https://www.ugr.es/~montero/matematicas/regresion_lineal.pdf

Moreno, F. (10 de Abril de 2015). *La importancia internacional de las lenguas*. Obtenido de
https://cervantesobservatorio.fas.harvard.edu/sites/default/files/010_informes_importancia_internacional_lenguas_0.pdf

Morgan, B. (16 de Diciembre de 2019). *100 estadísticas sobre transformación digital y experiencia del cliente*. Obtenido de
<https://www.forbes.com/sites/blakemorgan/2019/12/16/100-stats-on-digital-transformation-and-customer-experience/?sh=7929e11b3bf3>

Mundana. (15 de Octubre de 2016). *Los 6 beneficios que la formación académica aporta a la sociedad en la actualidad*. Obtenido de <https://www.mundana.us/blog/formacion-academica>

NETAPP. (23 de Octubre de 2023). *¿QUÉ ES LA INTELIGENCIA ARTIFICIAL?* Obtenido de
<https://www.netapp.com/es/artificial-intelligence/what-is-artificial-intelligence/>

- Páramo, O. (4 de Octubre de 2017). *La inteligencia artificial, muy lejos de la autoconciencia*.
Obtenido de https://unamglobal.unam.mx/global_revista/la-inteligencia-artificial-muy-lejos-de-la-autoconciencia/
- Pegasystems. (13 de Noviembre de 2023). *Empresa Autónoma*. Obtenido de <https://www.pega.com/es/technology/autonomous-enterprise>
- Perucha, C. (1 de Julio de 2022). *El Método de K Medias*. Obtenido de <https://uvadoc.uva.es/bitstream/handle/10324/58229/TFG-G5999.pdf?sequence=1>
- Pino, P. (1 de Noviembre de 2016). *Las tecnologías de hoy en un mundo globalizado*.
Obtenido de https://usbcali.edu.co/sites/default/files/9_tecnologias hoy.pdf
- Ponce, J., Torres, A., Quezada, F., & Silva, A. (1 de Marzo de 2014). *Inteligencia Artificial*.
Obtenido de https://www.researchgate.net/publication/269466259_Inteligencia_Artificial
- Pop, P. (27 de Enero de 2023). *¿Por qué hay tantas variaciones en el trabajo flexible?* Obtenido de <https://remote.pluria.co/es/variaciones-en-el-trabajo-flexible/>
- Ramírez, L. (5 de Enero de 2023). *Algoritmo k-means: ¿Qué es y cómo funciona?* Obtenido de <https://www.iebschool.com/blog/algoritmo-k-means-que-es-y-como-funciona-big-data/>
- Raya, A. (10 de Septiembre de 2021). *¿Qué es la globalización?* Obtenido de <https://elordenmundial.com/que-es-globalizacion/>
- Sandoval, L. (19 de Junio de 2018). *ALGORITMOS DE APRENDIZAJE AUTOMÁTICO PARA ANÁLISIS Y PREDICCIÓN DE DATOS*. Obtenido de http://www.redicces.org.sv/jspui/bitstream/10972/3626/1/Art6_RT2018.pdf
- Segovia, L. (1 de Enero de 2022). *Políticas públicas y gubernamentalidad del Sistema de Educación Superior del Ecuador en el periodo 2008-2018*. Obtenido de <https://repositorio.uasb.edu.ec/bitstream/10644/9042/1/T3962-MG-Segovia-Políticas.pdf>
- Seminarium Certificación. (10 de Noviembre de 2023). *Conozca la importancia de manejar idiomas en el mundo laboral*. Obtenido de <https://www.seminariumcertificacion.com/conozca-la-importancia-de-manejar-idiomas-en-el-mundo-laboral/#:~:text=Noticias->

,Conozca%20la%20importancia%20de%20manejar%20idiomas%20en%20el%20mundo
%20laboral,y%20genera%20diferencias%20de%20ingresos.

- TIBCO. (4 de Noviembre de 2023). *¿Qué es el aprendizaje no supervisado?* Obtenido de <https://www.tibco.com/es/reference-center/what-is-unsupervised-learning>
- Tokio School. (27 de Abril de 2022). *Conoce la historia del machine learning: ¡desde sus inicios!* Obtenido de <https://www.tokioschool.com/noticias/historia-machine-learning/>
- Uni Oviedo. (17 de enero de 2024). *El algoritmo k-means aplicado a clasificación y procesamiento de imágenes*. Obtenido de https://www.unioviedo.es/compnum/laboratorios_py/kmeans/kmeans.html
- UNIR. (22 de Abril de 2021). *Clustering: qué es y cuál es su uso en Big Data*. Obtenido de <https://www.unir.net/ingenieria/revista/clustering/>
- Universo Fórmulas. (16 de Enero de 2024). *La media cuadrática*. Obtenido de <https://www.universoformulas.com/estadistica/descriptiva/media-cuadratica/>
- UTEC. (4 de Diciembre de 2018). *Diferentes tipos de Inteligencia Artificial*. Obtenido de <https://educacion-ejecutiva.utec.edu.pe/blog/diferentes-tipos-inteligencia-artificial#:~:text=Teor%C3%ADa%20de%20la%20mente%3A%20Con,%2C%20deseos%2C%20necesidades%20y%20creencias.>
- Valera, C., Vásquez, R., & Iglesias, V. (12 de Marzo de 2009). *Comportamiento de abandono de la relación de un cliente con la empresa en un contexto de fallo y recuperación del servicio*. Obtenido de <https://digibuo.uniovi.es/dspace/bitstream/handle/10651/35608/1-s2.0-S1138575809700453-main.pdf?sequence=1>
- Vásquez, S. (19 de Diciembre de 2019). *¿Cómo aplicar machine learning en la opinión de los clientes?* Obtenido de <https://medium.com/@goodrebels/c%C3%B3mo-aplicar-machine-learning-en-la-opini%C3%B3n-de-los-clientes-3c64c2b17cdf>
- Velasquez, E. (10 de Septiembre de 2022). *Estado del Arte de Machine Learning y su Aplicación en el Experimento LHCb*. Obtenido de <https://mm.unah.edu.hn/dmsdocument/13661-estado-del-arte-de-machine-learning-y-su-aplicacion-en-el-experimento-lhcb-pdf>
- Vicerrectoría de Pregrado. (1 de Octubre de 2022). *Regresión Lineal Simple*. Obtenido de https://www.pace.otalca.cl/wp-content/uploads/2022/10/Sylvana_Regresionlinealsimple.pdf

- Westreicher, G. (1 de Junio de 2020). *Experiencia Laboral*. Obtenido de <https://economipedia.com/definiciones/experiencia-laboral.html>
- Winn, Z. (14 de Julio de 2023). *Un estudio encuentra que ChatGPT aumenta la productividad de los trabajadores para algunas tareas de escritura*. Obtenido de <https://news.mit.edu/2023/study-finds-chatgpt-boosts-worker-productivity-writing-0714>
- Zapata, B. (15 de Marzo de 2021). *Más universitarios consiguen empleo, según cifras del INEC; bachilleres se mantienen*. Obtenido de <https://www.eluniverso.com/noticias/economia/mas-universitarios-consiguen-empleo-segun-cifras-del-inec-bachilleres-se-mantienen-nota/>
- Zemsania. (24 de Noviembre de 2023). *Beneficios que aporta aplicar Machine learning en la empresa*. Obtenido de <https://zemsaniaglobalgroup.com/machine-learning-en-la-empresa/#:~:text=El%20Machine%20Learning%20permite%20analizar,Disminuci%C3%B3n%20de%20errores.>

ANEXOS

Anexo 1. Base de datos (study)

Funcion_Propietario	Inversion	Antiguedad	Nivel_educativo	Solvencia	Obligatoria
2	954,6192	0	1	16794	2
2	2231,07	0	1	12000	1
2	1784,856	0	1	12000	1
2	930,75372	0	1	16794	2
2	1829,6868	0	1	16794	2
2	2278,45726	0	1	16794	2
2	662,93	0	1	11800	1
2	499,75968	0	1	12000	1
2	1204,7778	1	1	12000	1
2	954,6192	0	1	16794	2
2	2007,963	0	1	12000	1
2	348,04692	0	1	12000	1
2	212,69534	0	1	12000	1
2	464,06256	0	1	12000	1
2	636,4128	0	1	16794	2
2	1312,6014	0	1	16794	2
2	1193,274	1	1	16794	2
2	2007,963	7	1	12000	1
2	1695,6132	7	1	12000	1
2	763,2	0	1	10000	1
2	428,4	0	1	10000	1
2	428,4	0	1	10000	1
2	345	0	1	10000	1
1	795,516	0	1	16794	2
1	1050,08112	0	1	16794	2
1	1050,08112	1	1	16794	2
1	1050,08112	1	1	16794	2
1	1145,54304	0	1	16794	2
1	1145,54304	0	2	16794	2
1	1236,38463	0	2	12000	1
1	1236,38463	23	2	12000	1
1	468,5247	0	2	12000	1
1	1500	0	2	10000	1
1	1952,18625	0	2	12000	1
1	1988,79	23	2	11800	1
1	2088,2295	10	2	11800	1
1	266,24102	0	2	12000	1
1	1896,4095	0	2	12000	1

1	531,73835	0	2	12000	1
1	392,66832	0	2	12000	1
3	535,4568	5	1	12000	1
3	2231,07	0	1	12000	1
3	2231,07	125	1	12000	1
3	468,5247	2	1	12000	1
2	1784,856	0	1	12000	1
2	1784,856	0	1	12000	1
2	964,56315	0	1	16794	2
2	964,56315	0	1	16794	2
2	670,22223	0	1	11800	1
2	2231,07	0	1	12000	1
2	859,15728	0	1	16794	2
2	1002,35016	0	1	16794	2
2	662,93	0	1	11800	1
2	662,93	0	1	11800	1
2	686,13255	0	1	11800	1
2	686,13255	0	1	11800	1
2	3882,0618	0	1	12000	1
2	1312,6014	0	1	16794	2
2	5533,0536	0	1	12000	1
2	2007,963	0	1	12000	1
2	2007,963	0	1	12000	1
2	1784,856	0	1	12000	1
2	1048,09233	0	1	16794	2
2	662,93	0	1	11800	1
2	662,93	0	1	11800	1
2	662,93	0	1	11800	1
2	3957,6921	0	1	16794	2
2	1174,04903	0	1	16794	2
2	1116	0	1	10000	1
2	612,54732	0	1	16794	2
2	1246,97133	0	1	16794	2
2	1508,16575	0	1	16794	2
2	2255,28786	0	1	16794	2
2	865,2	0	1	10000	1
2	1396,13058	0	1	16794	2
5	2311,96838	0	1	11800	1
2	1789,911	0	1	16794	2
2	1034,1708	0	1	16794	2
2	1182,66712	0	1	16794	2
2	1182,66712	0	1	16794	2
2	861,19302	0	1	12000	1
2	1050,08112	0	1	16794	2
2	1662,14715	0	1	12000	1
2	1050,08112	91	1	16794	2

2	1336,46688	6	1	16794	2
2	1017	0	1	10000	1
2	596,637	0	1	16794	2
2	356	0	1	10000	1
2	715,9644	0	1	16794	2
2	750,768225	1	1	11800	1
2	396	0	1	10000	1
2	445,48896	0	1	11800	1
2	696,0765	0	1	16794	2
2	819	0	1	10000	1
2	1020,9122	0	1	16794	2
2	584,70426	0	1	16794	2
2	267,7284	0	1	12000	1
2	535,4568	0	1	12000	1
2	1606,3704	0	1	12000	1
2	1020,9122	0	1	16794	2
2	435,80234	0	1	12000	1
2	1058	0	1	10000	1
2	300	0	1	10000	1
2	1020,9122	0	1	16794	2
2	1003,9815	0	1	12000	1
2	237,859284	0	1	11800	1
2	682,8179	0	1	16794	2
2	682,8179	0	1	16794	2
2	237,859284	0	1	11800	1
2	540	0	1	10000	1
2	456,3	0	1	10000	1
2	620,50248	0	1	16794	2
2	540	0	1	10000	1
2	1050,08112	0	1	16794	2
2	1710,487	6	1	12000	1
2	1710,487	0	1	12000	1
2	540	29	1	10000	1
2	540	0	1	10000	1
2	540	0	1	10000	1
2	540	0	1	10000	1
2	612	0	1	10000	1
2	1067,3173	0	1	16794	2
2	1933,594	0	1	12000	1
2	1710,487	0	1	12000	1
2	1008	0	1	10000	1
2	1673,3025	0	1	12000	1
2	1673,3025	4	1	12000	1
2	1606,3704	0	1	12000	1
2	1008	0	1	10000	1
2	811,42632	0	1	16794	2

2	725	0	1	10000	1
2	214,78932	0	1	11800	1
2	321,27408	0	1	12000	1
2	918,82098	0	1	16794	2
2	642,54816	0	1	12000	1
2	521	0	1	10000	1
2	811,42632	0	1	16794	2
2	1503,52524	0	1	16794	2
2	1486	0	1	10000	1
2	309,6	0	1	10000	1
2	1145,54304	0	1	16794	2
2	930,75372	0	1	16794	2
2	930,75372	2	1	16794	2
2	930,75372	0	1	16794	2
2	894,9555	0	1	16794	2
2	930,75372	0	1	16794	2
2	556,8612	0	1	16794	2
2	1002,35016	0	1	16794	2
2	1002,35016	0	1	16794	2
2	1060,688	0	1	16794	2
2	1784,856	0	1	12000	1
2	2275,6914	0	1	12000	1
2	2654,9733	0	1	12000	1
2	1002,35016	0	1	16794	2
2	1431,9288	0	1	16794	2
2	1431,9288	0	1	16794	2
2	558	0	1	10000	1
2	1050,08112	0	1	16794	2
2	662,93	0	1	11800	1
2	556,8612	0	1	16794	2
2	3882,0618	0	1	12000	1
2	994,395	0	1	16794	2
2	875,0676	0	1	16794	2
2	875,0676	0	1	16794	2
2	994,395	0	1	16794	2
2	936	0	1	10000	1
2	936	0	1	10000	1
2	1575,12168	0	1	16794	2
2	676,1886	0	1	16794	2
2	1784,856	0	1	12000	1
2	954,6192	0	1	16794	2
2	742,4816	0	1	16794	2
2	676,1886	0	1	16794	2
2	348,04692	0	1	12000	1
2	914,8434	0	1	16794	2
2	875,0676	0	1	16794	2

2	875,0676	0	1	16794	2
2	1193,274	0	1	16794	2
2	994,395	0	1	16794	2
2	1113,7224	0	1	16794	2
2	556,8612	0	1	16794	2
2	572,77152	0	1	16794	2
2	875,0676	0	1	16794	2
2	1050,08112	0	1	16794	2
2	596,637	0	1	16794	2
2	1575,12168	0	1	16794	2
2	875,0676	0	1	16794	2
2	572,77152	0	1	16794	2
2	875,0676	0	1	16794	2
2	1784,856	5	1	12000	1
2	1050,08112	0	1	16794	2
2	1044,11475	0	1	16794	2
2	1044,11475	0	1	16794	2
2	1044,11475	0	1	16794	2
2	1044,11475	0	1	16794	2
2	1044,11475	0	1	16794	2
2	1044,11475	0	1	16794	2
2	1044,11475	0	1	16794	2
2	1044,11475	0	1	16794	2
2	1784,856	0	1	12000	1
2	558	0	1	10000	1
2	1044,11475	0	1	16794	2
2	4004,77065	13	1	12000	1
2	464,06256	11	1	12000	1
2	982,8	0	1	10000	1
2	763,69536	0	1	16794	2
2	1073,9466	1	1	16794	2
2	787,56084	0	1	16794	2
2	819	0	1	10000	1
2	982,8	0	1	10000	1
2	572,77152	0	1	11800	1
2	572,77152	0	1	11800	1
2	702	0	1	10000	1
2	882	0	1	10000	1
2	1169,40852	0	1	16794	2
2	702	0	1	10000	1
2	397,8	0	1	10000	1
2	501,17508	0	1	16794	2
2	501,17508	0	1	16794	2
2	302,29608	0	1	16794	2
2	1097,81208	0	1	16794	2
2	540	0	1	10000	1
2	682,8179	0	1	16794	2

2	1896,4095	0	1	12000	1
2	1386,98185	0	1	12000	1
2	1386,98185	0	1	12000	1
2	1386,98185	0	1	12000	1
2	1075,4766	0	1	12002	1
2	715,9644	0	1	16794	2
2	1896,4095	0	1	12000	1
2	1365	0	1	10000	1
2	1365	0	1	10000	1
2	626,46885	0	1	16794	2
2	1312,6014	0	1	16794	2
2	2026,55525	0	1	12000	1
2	3709,52572	0	1	12000	1
2	892,30378	0	1	16794	2
2	449,4	0	1	10000	1
2	1217,13948	17	1	16794	2
2	294,50124	0	1	12000	1
2	392,66832	0	1	12000	1
2	811,42632	0	1	16794	2
2	1048,09233	0	1	16794	2
2	453,44412	1	1	11800	1
2	818,71855	0	1	11800	1
2	572,77152	0	1	11800	1
2	572,77152	0	1	11800	1
2	1784,856	0	1	12000	1
2	1145,54304	0	1	16794	2
2	270	0	1	10000	1
2	1443,86154	0	1	16794	2
2	3542,19547	0	1	12000	1
2	1372,1598	0	1	12002	1
2	563,4905	0	1	11800	1
2	1186,6447	0	1	16794	2
2	1186,6447	0	1	16794	2
2	984,45105	0	1	16794	2
2	984,45105	0	1	16794	2
2	715,9644	0	1	16794	2
2	453,44412	0	1	11800	1
2	534,98451	0	1	11800	1
2	662,93	0	1	11800	1
2	662,93	0	1	11800	1
2	954,6192	0	1	16794	2
2	1145,54304	0	1	16794	2
2	859,15728	0	1	16794	2
2	715,9644	0	1	16794	2
2	715,9644	0	1	16794	2
2	662,93	0	1	11800	1

2	954,6192	0	1	16794	2
2	662,93	0	1	11800	1
2	571,777125	0	1	11800	1
2	662,93	0	1	11800	1
2	715,9644	0	1	16794	2
2	715,9644	0	1	16794	2
2	964,56315	0	1	16794	2
2	662,93	0	1	11800	1
2	176,33938	0	1	11800	1
2	795,516	0	1	11800	1
2	285	0	1	10000	1
2	680,16618	0	1	11800	1
2	566,80515	0	1	11800	1
2	930,75372	0	1	16794	2
2	680,16618	0	1	11800	1
2	534,98451	0	1	11800	1
2	2208,7593	0	1	12000	1
2	712,826865	0	1	12000	1
2	662,93	0	1	11800	1
2	930,75372	0	1	16794	2
2	3882,0618	0	1	12000	1
2	589,00248	0	1	12000	1
2	1050,08112	0	1	16794	2
2	1252,9377	0	1	16794	2
2	1252,9377	0	1	16794	2
2	2275,6914	0	1	12000	1
2	3882,0618	0	1	12000	1
2	642,54816	0	1	12000	1
2	3765,30247	0	1	12000	1
2	428,36544	0	1	12000	1
2	428,36544	0	1	12000	1
2	321,27408	0	1	12000	1
2	662,93	0	1	11800	1
2	859,15728	0	1	16794	2
2	859,15728	0	1	16794	2
2	662,93	0	1	11800	1
2	859,15728	0	1	16794	2
2	662,93	0	1	11800	1
2	1717,9239	0	1	12000	1
2	1093,8345	0	1	16794	2
2	859,15728	0	1	16794	2
2	662,93	0	1	11800	1
2	662,93	0	1	11800	1
2	266,24102	0	1	12000	1
2	1193,274	0	1	16794	2
2	580,0782	0	1	12000	1

2	1784,856	0	1	12000	1
2	1044,14076	0	1	12000	1
2	696,09384	0	1	12000	1
2	1491,5925	0	1	16794	2
2	1093,8345	0	1	16794	2
2	1789,911	0	1	16794	2
2	157,11441	0	1	11800	1
2	859,15728	0	1	16794	2
2	1193,274	0	1	16794	2
2	1312,20507	0	1	12002	1
2	662,93	0	1	11800	1
2	1312,6014	0	1	16794	2
2	1312,6014	0	1	16794	2
2	660	0	1	10000	1
2	875,0676	0	1	16794	2
2	1252,9377	0	1	16794	2
2	662,93	0	1	11800	1
2	835,2918	0	1	16794	2
2	835,2918	0	1	16794	2
2	835,2918	1	1	16794	2
2	1252,9377	2	1	16794	2
2	1048,09233	2	1	16794	2
2	1252,9377	3	1	16794	2
2	1431,9288	0	1	16794	2
2	1431,9288	0	1	16794	2
2	662,93	0	1	11800	1
2	1050,08112	0	1	16794	2
2	1050,08112	0	1	16794	2
2	954,6192	0	1	16794	2
2	1431,9288	0	1	16794	2
2	1487,61492	0	1	16794	2
2	949,31576	0	1	16794	2
4	774	1	1	10000	1
4	774	1	1	10000	1
4	840,064896	2	1	16794	2
4	1183,35953	0	1	12000	1
4	596,637	0	1	11800	1

Anexo2. Script completo en RStudio

```
Study <- read.csv2("../DATA/DATOS EYM.csv")

#K MEDIAS PARA SEPARAR por grupos de importancia
install.packages("tidyverse")
library(tidyverse)
library(cluster)
library(factoextra)
install.packages("NbClust")
library(NbClust)

#Escalaamos
SKStudy<- scale(Study)

#Estimar el numero de cluster
NumClust<-NbClust(SKStudy, distance = "euclidean",
                  min.nc=2, max.nc=10, method = "kmeans",
                  index = "alllong")

#calculamos los dos clusters
k2 <- kmeans(SKStudy, centers = 2, nstart = 25)
k2
str(k2)

#plotear los cluster
fviz_cluster(k2, data = SKStudy)
fviz_cluster(k2, data = SKStudy, ellipse.type = "euclid",repe1 = TRUE,star.plot = TRUE) #ellipse.type= "t", "norm", "euclid"
fviz_cluster(k2, data = SKStudy, ellipse.type = "norm")

#Pasar Cluster a la base
Study$K_Means<-as.factor(k2$cluster)
table(Study$K_Means)

#Codifico la variable de respuesta
Study$K_Means <- factor(Study$K_Means,
                      levels = c(1,2),
                      labels = c("Estandar","Primium"))

Study$Funcion_Propietario <- factor(Study$Funcion_Propietario,
                                   levels = c(1,2,3,4,5),
                                   labels = c("Consultor Ecu ES","Consultor Ecu Idiomas",
                                              "Coordinador", "Director", "Operador"))

Study$Nivel_educativo <- factor(Study$Nivel_educativo,
                                levels = c(1,2),
                                labels = c("Bachillerato","Tercer nivel"))

#Creamos arbol
library(caret)

#para un arbol o cualquier derivado de un arbol de decision, se planta una semilla
set.seed(123)

#entrenamos los datos
dataentrenamiento <- createDataPartition(Study$K_Means, p=0.80, list = FALSE)

library(randomForest)

#Paso 4.- Armaamos el bosque aleatorio
RandomTreeModel <- randomForest(x=Study[dataentrenamiento, 1:6],
                                y=Study[dataentrenamiento,7],
                                ntree = 3000, keep.forest = TRUE)

RandomTreeModel$importance

#Paso 5.- Realizo la prediccion
Prediccion <- predict(RandomTreeModel,Study)
Prediccion
```

```

#Paso 6.- Armo la matriz de confusion
Matriz <- table(Study$K_Means, Prediccion, dnn = c("Actual", "Predicho"))
Matriz

#Valores clasificados
Study$Prediccion <- predict(RandomTreeModel,Study)

#Probabilidad de éxito
Study$RandProb <- predict(RandomTreeModel,Study, type = "prob")

#Tabla de frecuencia de tipo de cliente
Fre_Tipo <- as.data.frame(table(Study$Prediccion))

Tab_Fre_Tipo <- transform(Fre_Tipo,
                          FRel= round(prop.table(Fre_Tipo$Freq),3),
                          FreAcu = cumsum(Fre_Tipo$Freq),
                          FrReAc= cumsum(round(prop.table(Fre_Tipo$Freq),3)))

Tab_Fre_Tipo

G5 <- ggplot(Tab_Fre_Tipo, aes(x= Tab_Fre_Tipo$Var1,
                             y=Tab_Fre_Tipo$FRel))+
  geom_bar(stat = "identity", fill="#2E86C1", colour="black",size=0.5)+
  geom_text(aes(label=paste0(Tab_Fre_Tipo$FRel)),
            position= position_stack(vjust = .5))+
  labs(title = "Frecuencias de Tipo de cliente")

G5

#Grafica nivel educativo y tipo de cliente
G9 <- ggplot(Study, aes(Study$Nivel_educativo))+
  geom_bar(aes(fill=Study$Prediccion))+
  labs(title = "Tipo de cliente por nivel educativo ")
G9

```

```

> Study <- read.csv2("../Desktop/DATA/DATOS EYM (1).csv")
> #K MEDIAS PARA SEPARAR por grupos de importancia
> install.packages("tidyverse")
WARNING: Rtools is required to build R packages but is not currently installed. Please download and install the appropriate version of Rtools before proceeding:

https://cran.rstudio.com/bin/windows/Rtools/
Installing package into 'C:/Users/ximen/AppData/Local/R/win-library/4.3'
(as 'lib' is unspecified)
probando la URL 'https://cran.rstudio.com/bin/windows/contrib/4.3/tidyverse_2.0.0.zip'
Content type 'application/zip' length 430792 bytes (420 KB)
downloaded 420 KB

package 'tidyverse' successfully unpacked and MD5 sums checked

The downloaded binary packages are in
C:/Users/ximen/AppData/Local/Temp/RtmpuqP58v/downloaded_packages
> library(tidyverse)
— Attaching core tidyverse packages ————— tidyverse 2.0.0 —
✓ dplyr 1.1.4 ✓ readr 2.1.5
✓ forcats 1.0.0 ✓ stringr 1.5.1
✓ ggplot2 3.4.4 ✓ tibble 3.2.1
✓ lubridate 1.9.3 ✓ tidyr 1.3.0
✓ purrr 1.0.2
— Conflicts ————— tidyverse_conflicts() —
✖ dplyr::filter() masks stats::filter()
✖ dplyr::lag() masks stats::lag()
i Use the conflicted package to force all conflicts to become errors
> library(cluster)
> library(factoextra)
Welcome! want to learn more? See two factoextra-related books at https://goo.gl/ve3wBa
> install.packages("NbClust")
WARNING: Rtools is required to build R packages but is not currently installed. Please download and install the appropriate version of Rtools before proceeding:

https://cran.rstudio.com/bin/windows/Rtools/
Installing package into 'C:/Users/ximen/AppData/Local/R/win-library/4.3'
(as 'lib' is unspecified)
probando la URL 'https://cran.rstudio.com/bin/windows/contrib/4.3/NbClust_3.0.1.zip'
Content type 'application/zip' length 122888 bytes (120 KB)
downloaded 120 KB

```

```

package 'NbClust' successfully unpacked and MD5 sums checked

The downloaded binary packages are in
  C:\Users\ximen\AppData\Local\Temp\RtmpuqP58v\downloaded_packages
> library(NbClust)
> #Escalaamos
> SKStudy<- scale(Study)
> #Estimar el numero de cluster
> NumClust<-NbClust(SKStudy, distance = "euclidean",
+                   min.nc=2, max.nc=10, method = "kmeans",
+                   index = "alllong")
*** : The Hubert index is a graphical method of determining the number of clusters.
      In the plot of Hubert index, we seek a significant knee that corresponds to a
      significant increase of the value of the measure i.e the significant peak in Hubert
      index second differences plot.

*** : The D index is a graphical method of determining the number of clusters.
      In the plot of D index, we seek a significant knee (the significant peak in Dindex
      second differences plot) that corresponds to a significant increase of the value of
      the measure.

*****
* Among all indices:
* 6 proposed 2 as the best number of clusters
* 6 proposed 3 as the best number of clusters
* 1 proposed 4 as the best number of clusters
* 4 proposed 5 as the best number of clusters
* 1 proposed 6 as the best number of clusters
* 4 proposed 7 as the best number of clusters
* 5 proposed 10 as the best number of clusters

      ***** Conclusion *****

* According to the majority rule, the best number of clusters is 2

*****
There were 50 or more warnings (use warnings() to see the first 50)
> #calculamos los dos clusters
> k2 <- kmeans(SKStudy, centers = 2, nstart = 25)
> k2
# A tibble: 1 x 6
  cluster     size
  <dbl> <dbl>
1     2  186.0

K-means clustering with 2 clusters of sizes 186, 165

Cluster means:
  Funcion_Propietario  Inversion  Antiguedad Nivel_educativo Solvencia
1      0.05689788      0.04526150  0.03745744      0.1371192 -0.9167448
2     -0.06413943     -0.05102206 -0.04222475     -0.1545707  1.0334214

Obligatoria
1     -0.9405155
2      1.0602175

Clustering vector:
 [1] 2 1 1 2 2 2 1 1 1 2 1 1 1 2 2 2 1 1 1 1 1 2 2 2 1 1 1 1 1 2 2 2 2 2 2 1 1 1 1 1
 [35] 1 1 1 1 1 1 1 1 1 1 1 1 2 2 1 1 2 2 1 1 1 1 2 1 1 1 1 2 1 1 1 2 1 1 1 2 1 1 1 2 2
 [69] 1 2 2 2 2 1 2 1 2 2 2 2 2 1 2 1 2 2 1 2 1 1 1 2 1 2 2 2 1 1 2 1 1 2 1 1 1 2 1 1
[103] 1 2 1 1 2 2 1 1 1 2 1 2 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 2 1 1 1 2 1 1 1 2 1 1
[137] 2 2 1 1 2 2 2 2 2 2 2 2 2 1 1 1 2 2 2 1 2 1 2 1 2 2 2 2 2 2 2 2 1 1 2 2 1 1 2 2 1
[171] 2 2 2 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 1 2 2 2 2 2 2 2 2 2 2 1 1 2 1 1 1 1
[205] 2 2 2 1 1 1 1 1 1 2 1 1 2 2 2 2 1 2 1 1 1 1 1 2 1 1 1 2 2 1 1 1 2 2 1 1 2 1 2
[239] 1 1 2 2 1 1 1 1 1 1 2 1 2 1 1 1 2 2 2 2 2 1 1 1 1 2 2 2 2 2 2 2 2 1 2 1 1 1 1
[273] 2 2 2 1 1 1 1 1 1 2 1 1 1 1 1 2 1 1 2 2 1 1 1 1 1 1 1 1 1 1 1 1 2 2 2 1 2 1
[307] 1 2 2 1 1 1 2 1 1 1 1 2 2 2 1 2 2 1 1 2 2 1 1 2 2 1 2 2 1 2 2 2 2 2 2 2 2 2 2 1
[341] 2 2 2 2 2 2 1 1 2 1 1

within cluster sum of squares by cluster:
 [1] 1147.0052 260.3771
 (between_SS / total_SS = 33.0 %)

Available components:

 [1] "cluster"      "centers"      "totss"       "withinss"
 [5] "tot.withinss" "betweenss"    "size"        "iter"
 [9] "ifault"
> str(k2)
list of 9
 $ cluster      : int [1:351] 2 1 1 2 2 2 1 1 1 2 ...
 $ centers      : num [1:2, 1:6] 0.0569 -0.0641 0.0453 -0.051 0.0375 ...
 .. attr(*, "dimnames")=List of 2
 .. ..$ : chr [1:2] "1" "2"
 .. ..$ : chr [1:6] "Funcion_Propietario" "Inversion" "Antiguedad" "Nivel_educativo" ...
 $ totss       : num 2100

```

```

$ totss      : num 2100
$ withinss  : num [1:2] 1147 260
$ tot.withinss: num 1407
$ betweenss  : num 693
$ size       : int [1:2] 186 165
$ iter       : int 1
$ ifault     : int 0
- attr(*, "class")= chr "kmeans"
> #plotear los cluster
> fviz_cluster(k2, data = SKStudy)
> fviz_cluster(k2, data = SKStudy, ellipse.type = "euclid", repel = TRUE, star.plot = TRUE) #ellipse.type= "t", "norm", "euclid"
> #plotear los cluster
> fviz_cluster(k2, data = SKStudy)
> fviz_cluster(k2, data = SKStudy, ellipse.type = "euclid", repel = TRUE, star.plot = TRUE) #ellipse.type= "t", "norm", "euclid"
> fviz_cluster(k2, data = SKStudy, ellipse.type = "norm")
> #Pasar Cluster a la base
> Study$K_Means<-as.factor(k2$cluster)
> table(Study$K_Means)

```

```

 1  2
186 165
> #Codifico la variable de respuesta
> Study$K_Means <- factor(Study$K_Means,
+                       levels = c(1,2),
+                       labels = c("Estandar","Primum"))
> Study$Funcion_Propietario <- factor(Study$Funcion_Propietario,
+                                   levels = c(1,2,3,4,5),
+                                   labels = c("Consultor Ecu ES","Consultor Ecu Idiomas",
+                                             "Coordinador", "Director", "Operador"))
> Study$Nivel_educativo <- factor(Study$Nivel_educativo,
+                                levels = c(1,2),
+                                labels = c("Bachillerato","Tercer nivel"))
> library(caret)
Loading required package: lattice

```

Attaching package: 'caret'

The following object is masked from 'package:purrr':

lift

```

> #para un arbol o cualquier derivado de un arbol de decision, se planta una semilla
> set.seed(123)
> #entrenamos los datos
> dataentrenamiento <- createDataPartition(Study$K_Means, p=0.80, list = FALSE)
> library(randomForest)
randomForest 4.7-1.1
Type rfNews() to see new features/changes/bug fixes.

```

Attaching package: 'randomForest'

The following object is masked from 'package:dplyr':

combine

The following object is masked from 'package:gplot2':

margin

```

> #Paso 4.- Armamos el bosque aleatorio
> RandomTreeModel <- randomForest(x=Study[dataentrenamiento, 1:6],
+                                y=Study[dataentrenamiento,7],
+                                ntree = 3000, keep.forest = TRUE)
> RandomTreeModel$importance
      MeanDecreaseGini
Funcion_Propietario  0.4224111
Inversion            7.9514898
Antiguedad           0.4913578
Nivel_educativo     0.4734066
Solvencia            65.4976472
Obligatoria         64.1864229

```

```

> #Paso 5.- Realizo la prediccion
> Prediccion <- predict(RandomTreeModel,Study)
> Prediccion

```

1	2	3	4	5	6	7	8
Primum	Estandar	Estandar	Primum	Primum	Primum	Estandar	Estandar
9	10	11	12	13	14	15	16
Estandar	Primum	Estandar	Estandar	Estandar	Estandar	Primum	Primum
17	18	19	20	21	22	23	24
Primum	Estandar	Estandar	Estandar	Estandar	Estandar	Estandar	Primum
25	26	27	28	29	30	31	32

1	2	3	4	5	6	7	8
Prmium	Estandar	Estandar	Prmium	Prmium	Prmium	Estandar	Estandar
9	10	11	12	13	14	15	16
Estandar	Prmium	Estandar	Estandar	Estandar	Estandar	Prmium	Prmium
17	18	19	20	21	22	23	24
Prmium	Estandar	Estandar	Estandar	Estandar	Estandar	Estandar	Prmium
25	26	27	28	29	30	31	32
Prmium	Prmium	Prmium	Prmium	Prmium	Estandar	Estandar	Estandar
33	34	35	36	37	38	39	40
Estandar							
41	42	43	44	45	46	47	48
Estandar	Estandar	Estandar	Estandar	Estandar	Estandar	Prmium	Prmium
49	50	51	52	53	54	55	56
Estandar	Estandar	Prmium	Prmium	Estandar	Estandar	Estandar	Estandar
57	58	59	60	61	62	63	64
Estandar	Prmium	Estandar	Estandar	Estandar	Estandar	Prmium	Estandar
65	66	67	68	69	70	71	72
Estandar	Estandar	Prmium	Prmium	Estandar	Prmium	Prmium	Prmium
73	74	75	76	77	78	79	80
Prmium	Estandar	Prmium	Estandar	Prmium	Prmium	Prmium	Prmium
81	82	83	84	85	86	87	88
Estandar	Prmium	Estandar	Prmium	Prmium	Estandar	Prmium	Estandar
89	90	91	92	93	94	95	96
Prmium	Estandar	Estandar	Estandar	Prmium	Estandar	Prmium	Prmium
97	98	99	100	101	102	103	104
Estandar	Estandar	Estandar	Prmium	Estandar	Estandar	Estandar	Prmium
105	106	107	108	109	110	111	112
Estandar	Estandar	Prmium	Prmium	Estandar	Estandar	Estandar	Prmium
113	114	115	116	117	118	119	120
Estandar	Prmium	Estandar	Estandar	Estandar	Estandar	Estandar	Estandar
121	122	123	124	125	126	127	128
Estandar	Prmium	Estandar	Estandar	Estandar	Estandar	Estandar	Estandar
129	130	131	132	133	134	135	136
Estandar	Prmium	Estandar	Estandar	Estandar	Prmium	Estandar	Estandar
137	138	139	140	141	142	143	144
Prmium	Prmium	Estandar	Estandar	Prmium	Prmium	Prmium	Prmium
145	146	147	148	149	150	151	152
Prmium	Prmium	Prmium	Prmium	Prmium	Prmium	Estandar	Estandar
153	154	155	156	157	158	159	160
Estandar	Prmium	Prmium	Prmium	Estandar	Prmium	Estandar	Prmium
161	162	163	164	165	166	167	168
Estandar	Prmium	Prmium	Prmium	Prmium	Estandar	Estandar	Prmium
169	170	171	172	173	174	175	176
Prmium	Estandar	Prmium	Prmium	Prmium	Estandar	Prmium	Prmium
177	178	179	180	181	182	183	184
Prmium							
185	186	187	188	189	190	191	192
Prmium	Prmium	Prmium	Prmium	Prmium	Estandar	Prmium	Prmium
193	194	195	196	197	198	199	200
Prmium	Prmium	Prmium	Prmium	Prmium	Prmium	Estandar	Estandar
201	202	203	204	205	206	207	208
Prmium	Estandar	Estandar	Estandar	Prmium	Prmium	Prmium	Estandar
209	210	211	212	213	214	215	216
Estandar	Estandar	Estandar	Estandar	Estandar	Prmium	Estandar	Estandar
217	218	219	220	221	222	223	224
Prmium	Prmium	Prmium	Prmium	Estandar	Prmium	Estandar	Estandar
225	226	227	228	229	230	231	232
Estandar	Estandar	Estandar	Prmium	Estandar	Estandar	Estandar	Prmium
233	234	235	236	237	238	239	240
Prmium	Estandar	Estandar	Prmium	Estandar	Prmium	Estandar	Estandar
241	242	243	244	245	246	247	248
Prmium	Prmium	Estandar	Estandar	Estandar	Estandar	Estandar	Prmium
249	250	251	252	253	254	255	256
Estandar	Prmium	Estandar	Estandar	Estandar	Prmium	Prmium	Prmium
257	258	259	260	261	262	263	264
Prmium	Prmium	Estandar	Estandar	Estandar	Estandar	Prmium	Prmium
265	266	267	268	269	270	271	272
Prmium	Prmium	Prmium	Estandar	Prmium	Estandar	Estandar	Estandar
273	274	275	276	277	278	279	280
Prmium	Prmium	Prmium	Estandar	Estandar	Estandar	Estandar	Estandar
281	282	283	284	285	286	287	288
Estandar	Prmium	Estandar	Estandar	Estandar	Estandar	Estandar	Prmium
289	290	291	292	293	294	295	296
Estandar	Estandar	Prmium	Prmium	Prmium	Estandar	Estandar	Estandar
297	298	299	300	301	302	303	304
Estandar	Estandar	Estandar	Estandar	Estandar	Prmium	Prmium	Estandar
305	306	307	308	309	310	311	312
Prmium	Estandar	Estandar	Prmium	Prmium	Estandar	Estandar	Estandar
313	314	315	316	317	318	319	320
Prmium	Estandar	Estandar	Estandar	Estandar	Prmium	Prmium	Prmium
321	322	323	324	325	326	327	328
Estandar	Prmium	Prmium	Estandar	Estandar	Prmium	Prmium	Estandar
329	330	331	332	333	334	335	336
Prmium	Prmium	Estandar	Prmium	Prmium	Prmium	Prmium	Prmium

```

329      330      331      332      333      334      335      336
Prmium Prmium Estandar Prmium Prmium Prmium Prmium Prmium
337      338      339      340      341      342      343      344
Prmium Prmium Prmium Estandar Prmium Prmium Prmium Prmium
345      346      347      348      349      350      351
Prmium Prmium Estandar Estandar Prmium Estandar Estandar
Levels: Estandar Prmium
> #Paso 6.- Armo la matriz de confusion
> Matriz <- table(Study$K_Means, Prediccion, dnn = c("Actual", "Predicho"))
> Matriz
      Predicho
Actual Estandar Prmium
Estandar 186      0
Prmium    0     165
> Matriz
      Predicho
Actual Estandar Prmium
Estandar 186      0
Prmium    0     165
> Matriz
      Predicho
Actual Estandar Prmium
Estandar 186      0
Prmium    0     165
> #Valores clasificados
> Study$Prediccion <- predict(RandomTreeModel,Study)
> #Valores clasificados
> Study$Prediccion <- predict(RandomTreeModel,Study)
> #Probabilidad de exito
> Study$RandProb <- predict(RandomTreeModel,Study, type = "prob")
> #Tabla de frecuencia de tipo de cliente
> Fre_Tipo <- as.data.frame(table(Study$Prediccion))
> Tab_Fre_Tipo <- transform(Fre_Tipo,
+                           FReI= round(prop.table(Fre_Tipo$Freq),3),
+                           FReAcu = cumsum(Fre_Tipo$Freq),
+                           FrReAc= cumsum(round(prop.table(Fre_Tipo$Freq),3)))
> Tab_Fre_Tipo

> Tab_Fre_Tipo
  Var1 Freq FReI FReAcu FrReAc
1 Estandar 186 0.53 186 0.53
2 Prmium 165 0.47 351 1.00
> G5 <- ggplot(Tab_Fre_Tipo, aes(x= Tab_Fre_Tipo$Var1,
+                               y=Tab_Fre_Tipo$FReI))+
+   geom_bar(stat = "identity", fill="#2E86C1", colour="black",size=0.5)+
+   geom_text(aes(label=paste0(Tab_Fre_Tipo$FReI)),
+             position= position_stack(vjust = .5))+
+   labs(title = "Frecuencias de Tipo de cliente")
Warning message:
Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
Please use `linewidth` instead.
This warning is displayed once every 8 hours.
Call `lifecycle::last_lifecycle_warnings()` to see where this warning
was generated.
> G5
Warning messages:
1: Use of `Tab_Fre_Tipo$Var1` is discouraged.
i Use `Var1` instead.
2: Use of `Tab_Fre_Tipo$FReI` is discouraged.
i Use `FReI` instead.
3: Use of `Tab_Fre_Tipo$FReI` is discouraged.
i Use `FReI` instead.
4: Use of `Tab_Fre_Tipo$Var1` is discouraged.
i Use `Var1` instead.
5: Use of `Tab_Fre_Tipo$FReI` is discouraged.
i Use `FReI` instead.
> G9 <- ggplot(Study, aes(Study$Nivel_educativo))+
+   geom_bar(aes(fill=Study$Prediccion))+
+   labs(title = "Tipo de cliente por nivel educativo ")
> G9
Warning message:

```



DECLARACIÓN Y AUTORIZACIÓN

Nosotros, **Cuesta Asanza, Ximena Margarita** con C.C: # 0923733745 y **Duran Encalada, Eduardo Javier**, con C.C # 0705299824 autores del trabajo de titulación curricular: **Creación de un modelo clasificador para oportunidades de captación en el exterior por Machine Learning de la empresa EDEX Education Group**, previo a la obtención del título de **Licenciado en Negocios Internacionales** en la Universidad Católica de Santiago de Guayaquil.

1.- Declaramos tener pleno conocimiento de la obligación que tienen las instituciones de educación superior, de conformidad con el Artículo 144 de la Ley Orgánica de Educación Superior, de entregar a la SENESCYT en formato digital una copia del referido trabajo de titulación para que sea integrado al Sistema Nacional de Información de la Educación Superior del Ecuador para su difusión pública respetando los derechos de autor.

2.- Autorizamos a la SENESCYT a tener una copia del referido trabajo de titulación, con el propósito de generar un repositorio que democratice la información, respetando las políticas de propiedad intelectual vigentes.

Guayaquil, 16 de febrero del 2024

f. _____

Cuesta Asanza, Ximena Margarita

C.C: 0923733745

f. _____

Duran Encalada, Eduardo Javier

CC: 0705299824

REPOSITORIO NACIONAL EN CIENCIA Y TECNOLOGÍA

FICHA DE REGISTRO DE TESIS/TRABAJO DE TITULACIÓN

TEMA Y SUBTEMA:	Creación de un modelo clasificador para oportunidades de captación en el exterior por Machine Learning de la empresa EDEX Education Group.		
AUTOR(ES)	Cuesta Asanza, Ximena Margarita Duran Encalada, Eduardo Javier		
REVISOR(ES)/TUTOR(ES)	Ing. Carrera Buri, Félix Miguel, Mgs.		
INSTITUCIÓN:	Universidad Católica de Santiago de Guayaquil		
FACULTAD:	Facultad de Economía y Empresa		
CARRERA:	Negocios Internacionales		
TÍTULO OBTENIDO:	Licenciado en Negocios Internacionales		
FECHA DE PUBLICACIÓN:	16 de febrero del 2024	No. DE PÁGINAS:	76
ÁREAS TEMÁTICAS:	Economía, Innovaciones, Negocios		
PALABRAS CLAVES/ KEYWORDS:	Machine Learning, estudios en el extranjero, globalización, RandomForest, RStudio		
RESUMEN	<p>Este trabajo tiene como objetivo crear un modelo clasificador para la captación de clientes en el exterior de la empresa EDEX Education Group basado en Machine Learning para identificar y evaluar las oportunidades de captación en el ámbito internacional, EDEX Education Group es una empresa que se dedica a la consultoría de estudios en el extranjero, cuenta con oficinas en distintos países pero debido a la pandemia laboran en algunos países a través de video llamadas, en Ecuador cuenta con oficinas las ciudades de Quito, Machala y Guayaquil. La globalización nos ha brindado la oportunidad de poder tomar cursos de idiomas, certificaciones y carreras universitarias en cualquier país del mundo y este proyecto es un enfoque que puede ayudar a mejorar significativamente la eficiencia del proceso de captación y así tomar decisiones mas exactas a la hora de trabajar con cada cliente y sus características a través de las consultorías personalizadas, empleando el algoritmo de RandomForest en RStudio.</p>		
ADJUNTO PDF:	<input checked="" type="checkbox"/> SI	<input type="checkbox"/> NO	
CONTACTO CON AUTOR/ES:	Teléfono: +59399222 5502 +59399847541245	E-mail: Ximena Margarita Cuesta Asanza ximenacuesta2001@gmail.com	
CONTACTO CON LA INSTITUCIÓN (COORDINADOR DEL PROCESO UTE)::	Nombre: Freire Quintero, Cesar Enrique Teléfono: +593 99 009 0702 E-mail: cesar.freire@cu.ucsg.edu.ec		
SECCIÓN PARA USO DE BIBLIOTECA			
Nº. DE REGISTRO (en base a datos):			
Nº. DE CLASIFICACIÓN:			
DIRECCIÓN URL (tesis en la web):			